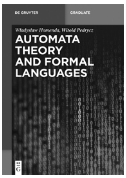
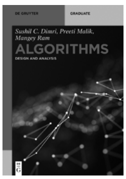
Mou-Hsiung Chang

Theory of Quantum Information with Memory

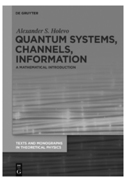
Also of Interest



Automata Theory and Formal Languages
Wladyslaw Homenda, Witold Pedrycz, 2022
ISBN 978-3-11-075227-4, e-ISBN (PDF) 978-3-11-075230-4,
e-ISBN (EPUB) 978-3-11-075231-1



Algorithms
Design and Analysis
Sushil C. Dimri, Preeti Malik, Mangey Ram, 2021
ISBN 978-3-11-069341-6, e-ISBN (PDF) 978-3-11-069360-7,
e-ISBN (EPUB) 978-3-11-069375-1



Quantum Systems, Channels, Information
A Mathematical Introduction
Alexander S. Holevo, 2019
ISBN 978-3-11-064224-7, e-ISBN (PDF) 978-3-11-064249-0,
e-ISBN (EPUB) 978-3-11-064240-7

Mou-Hsiung Chang

Theory of Quantum Information with Memory

DE GRUYTER

Author
Ph. D. Mou-Hsiung Chang
7709 Silver View Ln
Raleigh, NC 27613
USA

ISBN 978-3-11-078799-3
e-ISBN (PDF) 978-3-11-078810-5
e-ISBN (EPUB) 978-3-11-078819-8

Library of Congress Control Number: 2022939327

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Cover image: sakkmesterke / iStock / Getty Images Plus
Typesetting: VTeX UAB, Lithuania
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Preface

Quantum information theory has now become a significant and important branch of scientific research. This is evidenced by its research activities during the last decade that have resulted in a huge body of literatures. However, a vast majority of these works have only focused on finite-dimensional and memoryless quantum channels without addressing increasingly important and yet realistic issues of infinite dimensionality and memory effects in quantum communication. This monograph is the first in a book form that addresses these two issues. It aims to give a systematic and mathematically rigorous treatment and to address the mathematical challenges that these two issues have posed.

This book is the product of a set of research notes that the author prepared during his study, learning and research on these two issues. Its content is largely based on the current account of relevant research results, published or not-yet-published, and contributed by many prominent researchers. The list of bibliography is certainly not exhaustive and is likely to have omitted works done by many researchers. The author apologizes for any inadvertent omissions of their works.

This book is intended to be as self-contained as possible in terms of mathematical prerequisites. With the exception of a few well-known theorems in functional analysis and operator theory that are stated in Chapter 1 without proofs, detailed proofs are provided for most of all other results for the reader's convenience and for reader's preparations in order to familiarize themselves with mathematical and theoretical aspects of the quantum information. On the other hand, mathematical topics are presented only if they will be used either as tools for rigorously treating relevant topics in quantum information or they will provide insights and motivations for extending the frontier in the field.

The birth of this book would not have been possible without the contributions of the following groups of people. The author would first like to thank many prominent researchers in this subject area who have made tremendous contributions in the literature. Their research and publications have greatly influenced the style and the presentation of many topics in this book. The author would also like to thank the readers of the earlier draft of this book for their suggestions that have led to this much improved final version. Last but not least, the author would especially like to thank Dr. Damiano Sacco of De Gruyter for his foresight and enthusiasm on this book project and Karolina Sobanska for her editorial assistance during and throughout publication of this project.

Raleigh, North Carolina, USA, April 2022 Mou-Hsiung Chang

https://doi.org/10.1515/9783110788105-201

Introduction

One of the most common and essential tasks of everyday life is transmission or communication of information. Today's society is in continuous and permanent communication: we often exchange messages via a classical communication device such as telephones, mobile phones, the internet, the radio and in many other ways. Conventionally communications through classical devices, messages are encoded in a sequence of bits. In this case, the signals, which carry the information corresponding to voltages or strong light pulses, the physical device is usually modeled by a *classical channel*. In modern day's communication, we also often communicate through quantum devices. Quantum messages as well as classical messages can both be transmitted via quantum communications devices, which incorporate intrinsically quantum mechanics effects.

Traditionally, two main approaches to quantum information processing and quantum communication have been pursued. On one hand, a "digital" one, according to which information, is encoded in systems with a discrete quantum unit (such as qubits or qudits) with a finite number of degrees of freedom. Typical examples of qubit implementations are the nuclear spins of individual atoms in a molecule, the polarization of photons, ground/excited states of trapped ions, etc. In parallel, an "analog" approach has also been devised, based on quantum information and correlations being encoded in degrees of freedom with a continuous spectrum (continuous variables). These are often associated with positions and momentum of quantized particles. This second approach has witnessed considerable success due to its versatility, with implementations often encompassing different physical systems, e. g., light quadratures and collective magnetic moments of atomic ensembles, which obey the same canonical algebra. In either case, laws of quantum mechanics dictate the behavior of the signals and the underlying device is modeled by a quantum channel.

In both classical and quantum communication scenarios, an immediate question arises: what is the maximal amount of classical information in the unit of bits (for classical communication) or of quantum information in the basic unit of qubits or qudits (for quantum communication) that can be transmitted reliably per channel use? In other words, what is the capacity of the channel? The answer to this question is quantified by the *classical capacity* of the classical channel and the classical capacity and quantum capacity of quantum channels in modern day communication depending on whether classical or quantum information is transmitted via the quantum channel and what resources are used in the transmission.

The classical information theory developed by Claude Shannon in his 1948 seminal paper [140] centered around the investigation of the classical capacity of classical channels and achievability of channel capacity (see also Cover and Thomas [29]). Encouraged by the success of classical information theory and the need to efficiently apply quantum resources for modern day information processing and communication, the creation of quantum information theory began with pioneer works by Holevo [67, 68] and continued to grow for the last couple of decades (see the monographs by

https://doi.org/10.1515/9783110788105-202

Holevo [77], Hayashi [61], Ohya and Petz [121], Watrous [173] and Wilde [178] for systematic expositions of some topics in mostly finite- dimensional theory).

What is this book all about?

Quantum information theory has now become a significant branch of research during the last few years. However, most of these works have focused on finite-dimensional and memoryless quantum channels without addressing increasingly important issues of infinite dimensionality and memory effects on quantum communication. The purpose of this book is to give a systematic and mathematically rigorous treatment of these two important and yet realistic issues and to address the mathematical challenges these two issues have posed.

A. The issues of infinite dimensionality

Why infinite-dimensional quantum information theory? Recently, an important class of infinite-dimensional Gaussian channels (see, e. g., Holevo and Werner [86] and Giovanneti et al. [54]) have been discovered and created in laboratories that can be immediately implemented in communication. Although many questions for Gaussian Bosonic systems with a finite number of modes can be solved with finite-dimensional matrix techniques, a general underlying Hilbert space operator analysis is indispensable. Moreover, it was observed recently by Shirokov [141, 142] that Shor's proof of global equivalence of different forms of the famous additivity conjecture is related to weird discontinuity of the χ-capacity in the infinite-dimensional case. All of this calls for a mathematically rigorous treatment involving specific results from the infinite-dimensional operator theory in a Hilbert space and measure theory.

In addition, it is well known that many relevant quantum properties behave differently in infinite-dimensional spaces than those in finite-dimensional spaces. For example, the properties of the entropy for infinite- and finite- dimensional Hilbert spaces differ quite substantially. In the latter case, the entropy is a bounded continuous function on the space of quantum states, while in the former it is lower semicontinuous but discontinuous at every point, and infinite most everywhere in the sense that the set of states with finite entropy is a first category subset of the collection of quantum states (see, e. g., Wehrl [175]).

There are two important features essential for channels in infinite dimensions. One is the necessity of the input constraints (such as the mean energy constraint for Gaussian channels) to prevent channels from having infinite capacities, although considering input constraints was recently shown quite useful also in the study of the additivity conjecture for channels in finite dimensions. Another is the natural appearance of infinite and, in general, continuous state ensembles understood as probability measures on the set of all quantum states. By using compactness criteria from probability theory and operator theory, one can show that the set of all generalized ensembles with the average in a compact set of states is itself a compact subset of the set of all probability. Interesting and important features in infinite-dimensional quantum channels such as χ-capacity have been demonstrated in Holevo and Shirokov [80, 82] and Lindblad [107]. In [158], Shirokov and Holevo developed an approximation approach to infinite-dimensional quantum channels based on a detailed investigation of continuity properties of entropic characteristics of quantum channels and operations. Recently, mutual information and coherent information for infinite-dimensional quantum channels are established and discussed by Holevo and Shirokov in [82].

The research works and rationales cited above call for a systematic treatment of infinite-dimensional quantum information theory, which is emphasized in the book.

B. The issues of channel with memory

Memory effects have appeared in nonquantum systems (such as classical and stochastic systems) as well as quantum systems. Major research efforts devoted to nonquantum systems have resulted in many important breakthroughs (see the monograph by Chang [21] and references contained therein). However, research on memory effects in quantum communication has just begun in recent years.

Why quantum channel with memory? We now give rationales for emphasizing systematic and mathematical treatment of memory effects on quantum channels in this book. First, the vast majority of the work on quantum communication has been concerned with the study of memoryless configurations on quantum channels where sequences of exchanged quantum carriers are supposed to undergo the action of noisy transformations, which affect them independently and identically. In this scenario, coding theorems have been derived, which allow one to express the various capacities of the communication line in terms of rather compact entropic formulas. For instance, the classical (resp., quantum) capacity of a memoryless quantum channel is characterized in terms of the Holevo (resp., coherent) information (see Barnum, Nielsen and Schumacher [5], Devetak [37], Holevo [69], Lloyd [109], Schumacher and Nielsen, [162], Schumacher and Westmoreland [163], Shor [159]). The memoryless assumption is indeed a useful hypothesis, which simplifies the input–output mapping induced by the noise. However, in real communication, assumption of independent actions of multiple uses of the memoryless channel cannot be justified. For instance, with increasing signal feeding rates, successive transmissions happen so rapidly that the environment may retain a "memory" of past events, Similarly, in quantum information processors, especially in solid-state implementations, qubits may be so closely spaced that the same environmental degree of freedom will interact jointly with several of them (even if they are not nearest neighbors) leading to cross-talks and correlations in the noise (see, e. g., Duan and Guo [43]). The consideration of spatial and temporal memory effects is therefore becoming increasingly pressing with the continuing miniaturization of information processing devices and with increasing communication rates through channels. Moreover, from a fundamental point of view, quantum memory channels provide a general framework, which encompasses the memoryless ones as a special case. Apparently, the interest toward information transmission through quantum channels with memory spread after a model introduced by Macchiavello and Palma [111] and intensive research on memory quantum channels is followed and summarized in the review paper by Caruso et al. [18].

This book consists of a total of sixteen chapters, all of which are in an infinite-dimensional setting and have never been covered elsewhere in book forms. The last three chapters are devoted to memory channels of various types. While memory effects on quantum channels is one of the main emphases in this book, they are placed at the end of the book because memory effects can physically take place only when channels are used repeatedly and these topics can be presented only after the prerequisite material is introduced and treated.

While Gaussian states and Gaussian channels provide ample real world examples of infinite-dimensional theory of quantum information, they are not discussed here at all due to page limitations of an already lengthy volume of this book. The theory of Gaussian quantum information constitutes substantial research efforts documented in the literature and it definitely deserves more serious and thorough treatment in a separate book by itself. This book also put disproportional emphasis on infinite-dimensional classical capacities of various types of quantum channels instead of quantum capacities. This is because, while some results on finite-dimensional quantum capacities have been obtained, developments on infinite-dimensional quantum capacities are still in their infancy. Concerted research efforts are much needed in order to achieve significant progress in this important area for modern quantum communications.

Contents

Preface —— V

Introduction —— VII

1 Basic notation and preliminaries —— 1
1.1 Complex Hilbert and Banach spaces —— **1**
1.2 Linear operators and their adjoints —— **7**
1.2.1 Adjoint and self-adjoint operators —— **8**
1.2.2 Bounded linear operators —— **9**
1.3 Positive operators —— **13**
1.4 Resolvent set and spectrum —— **15**
1.5 Unitary operators —— **16**
1.6 Projection operators —— **17**
1.7 Finite rank and compact linear operators —— **21**
1.7.1 Finite rank linear operators —— **21**
1.7.2 Compact linear operators —— **22**
1.8 Trace-class and Hilbert–Schmidt operators —— **25**
1.8.1 Trace-class operators —— **25**
1.8.2 Hilbert–Schmidt operator —— **29**

2 Formulation of quantum systems —— 33
2.1 Operator topologies —— **34**
2.2 Operator algebras —— **36**
2.2.1 C^*-algebras —— **37**
2.2.2 von Neumann algebras —— **39**
2.3 Bounded linear functionals —— **42**
2.3.1 Bounded linear functionals on $\mathbb{H}$ —— **44**
2.3.2 $\mathcal{A}^*$ and $\mathcal{A}_*$ —— **48**
2.4 Quantum states —— **53**
2.5 GNS representation —— **56**
2.6 Quantum observables and measurements —— **58**
2.6.1 Positive operator valued measures —— **58**
2.6.2 Quantum observables —— **59**
2.6.3 Quantum measurements —— **60**
2.7 Tensor products and direct sums —— **62**
2.7.1 Tensor products of Hilbert spaces and operators —— **62**
2.7.2 Direct sum of Hilbert spaces and operators —— **65**
2.8 Composite quantum systems —— **67**
2.8.1 Composite system as tensor products —— **67**
2.8.2 Partial traces —— **67**

3 **Probability measures and convex functions on $\mathcal{S}(\mathbb{H})$ — 71**
3.1 Probability measures on $\mathcal{S}(\mathbb{H})$ — **71**
3.1.1 Support of Borel measures — **71**
3.2 Some compactness criteria — **74**
3.2.1 Compactness criteria on $\mathcal{S}(\mathbb{H})$ — **74**
3.2.2 Compactness criteria on $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ — **79**
3.3 Barycenters — **80**
3.3.1 A brief review of convex analysis — **80**
3.3.2 Properties of barycenter — **81**
3.4 Convex functions on $\mathcal{S}(\mathbb{H})$ — **88**
3.4.1 σ-convexity and μ-convexity — **88**
3.4.2 Convex hull and convex closure — **91**

4 **Completely positive maps — 95**
4.1 Definitions and properties — **95**
4.2 Some technical results — **102**
4.3 Stinespring representation — **108**
4.4 Kraus representation — **113**

5 **Quantum channels and operations — 117**
5.1 Quantum operations — **117**
5.2 Quantum channels — **118**
5.2.1 Quantum channels in the Schrodinger picture — **119**
5.2.2 Quantum channels in the Heisenberg picture — **120**
5.3 Representations and dilations of channels — **123**
5.3.1 Stinespring representation — **123**
5.3.2 Unitary dilation — **124**
5.3.3 Kraus representation — **127**
5.4 Structural properties of quantum channels — **128**
5.4.1 Convexity of channels — **128**
5.4.2 Concatenation of channels — **128**
5.4.3 Tensor product of channels — **129**
5.5 Purification — **129**
5.6 Reversible quantum channels — **130**
5.7 Complementary channels — **132**

6 **Approximation and convergence of quantum channels — 133**
6.1 Distinguishability of quantum states — **133**
6.1.1 Fidelity of quantum states — **133**
6.1.2 Bures distance — **145**
6.2 Channel fidelity — **145**
6.2.1 Monotonicity of channel fidelity — **146**

6.2.2 Ensemble average and entanglement fidelities —— **146**
6.3 Norms on unconstrained channels —— **149**
6.3.1 Diamond norm —— **149**
6.3.2 Complete boundedness norm —— **149**
6.3.3 Bures distance —— **151**
6.4 Norms on constrained channels —— **153**
6.4.1 Energy-constrained operator norm —— **154**
6.4.2 Constrained diamond norms and Bures distances —— **156**
6.5 Approximations of quantum channels —— **159**
6.6 Convergences of channels —— **161**
6.6.1 Strong and uniform convergence —— **161**
6.6.2 Strong* convergence of channels —— **165**

7 von Neumann entropy —— 169
7.1 von Neumann entropy on $\mathcal{S}(\mathbb{H})$ —— **169**
7.2 Properties of $H(\cdot)$ on $\mathcal{S}(\mathbb{H})$ —— **171**
7.3 $H(\cdot)$ on subsets of $\mathcal{S}(\mathbb{H})$ —— **177**
7.3.1 Closed convex subsets $\mathcal{A} \subset \mathcal{S}(\mathbb{H})$ —— **177**
7.3.2 $\mathcal{K}_{\mathbf{H}}(h)$ —— **178**
7.3.3 $H(\cdot)$ on an arbitrary closed subset $\mathcal{K}$ —— **192**
7.4 $H(\cdot)$ extended to $\mathfrak{T}_+(\mathbb{H})$ —— **196**
7.4.1 Properties of $H(\cdot)$ on $\mathfrak{T}_+(\mathbb{H})$ —— **197**
7.4.2 Discontinuity of extended quantum entropies —— **200**

8 Relative and conditional entropies —— 203
8.1 Quantum relative entropy —— **203**
8.2 $H(\cdot\|\cdot)$ extended to $\mathfrak{T}_+(\mathbb{H}) \times \mathfrak{T}_+(\mathbb{H})$ —— **210**
8.3 Quantum conditional entropy —— **213**
8.3.1 Definitions —— **213**
8.3.2 Properties of conditional entropy —— **214**

9 Channel output entropies —— 223
9.1 Monotonicity of output relative entropy and Petz’s theorem —— **223**
9.1.1 Monotonicity of output relative entropy —— **223**
9.1.2 Petz’s theorem on output relative entropy —— **225**
9.2 Continuity of output entropies —— **228**
9.2.1 Positive linear map Φ —— **228**
9.2.2 Quantum operations —— **234**
9.3 Convex closure of output entropy —— **239**

10 Quantum entanglement —— 243
10.1 Schmidt decomposability of states —— **243**

10.2 Separability of states — 250
10.2.1 Matrix of amplitudes — 250
10.2.2 Separability — 255
10.3 Measurements of entanglement — 263
10.3.1 Entropy of entanglement — 266
10.3.2 Relative entropy of entanglement — 269
10.3.3 Entanglement of formation — 272

11 Quantum mutual and coherent information — 277
11.1 Quantum mutual information — 277
11.1.1 Finite-dimensional case — 284
11.1.2 Infinite-dimensional case — 286
11.2 Coherent information — 288

12 Holevo χ-capacity — 295
12.1 The χ-functions — 295
12.1.1 Input χ-function — 295
12.1.2 Output χ-function — 296
12.2 Holevo χ-capacities — 303
12.2.1 Optimal ensembles — 305
12.2.2 Compact constraint $\mathcal{A}$ — 312
12.2.3 Convex constraint $\mathcal{A}$ — 313
12.3 Continuity of Holevo χ-capacity — 320
12.4 Holevo χ-quantity and coherent information — 324
12.5 Additivity of Holevo χ-capacities — 327
12.5.1 Hastings' counterexamples to additivity — 328
12.5.2 Additivity of unconstrained χ-quantity — 329
12.5.3 Tensor product of constrained channels — 334
12.5.4 Additivity of constrained Holevo χ-capacity — 335

13 Classical capacities of memoryless channels — 339
13.1 Transmissions of classical information — 341
13.1.1 Preparation of classical information — 341
13.1.2 Encoding and decoding of classical information — 342
13.2 Unconstrained classical capacity — 346
13.2.1 Holevo–Schumacher–Westmoreland theorem — 347
13.2.2 Quantum version of Feinstein's lemma — 348
13.3 Classical capacity of constrained channels — 355
13.4 Entanglement-assisted classical capacity — 358
13.4.1 Finite-dimensional channels — 360
13.4.2 Constrained infinite-dimensional channels — 367
13.4.3 Continuity of $C_{\mathrm{ea}}(\cdot) : \mathfrak{QC}(A, B) \to [0, +\infty]$ — 374

13.5 Comparison of classical capacities — 377
13.5.1 Finite-dimensional unconstrained capacities — 377
13.5.2 Infinite-dimensional constrained capacities — 379

14 Structure of quantum memory channels — 383
14.1 Representations of memoryless channels — 384
14.1.1 Kraus representation in the Schrodinger picture — 384
14.1.2 Kraus representation in the Heisenberg picture — 385
14.1.3 Unitary representation in the Schrodinger picture — 385
14.2 Constructive approach to memory channels — 386
14.2.1 Unitary representation in the Schrodinger picture — 389
14.2.2 Unitary representation in the Heisenberg picture — 390
14.2.3 Kraus representation of the memory channel — 391
14.3 Quasilocal algebras approach — 391
14.3.1 Construction of quasilocal algebras — 391
14.3.2 Structure of causal channels — 394
14.4 Forgetful channels — 397
14.4.1 Definition of forgetful channels — 397
14.4.2 Topological properties — 404

15 Channels with Markovian memory — 405
15.1 A brief review on Markov chains — 405
15.2 Constructions of Markov memory models — 407
15.3 Channels with ergodic Markovian memory — 409
15.3.1 Classical capacity — 409
15.4 Channels with general Markov memory — 422
15.4.1 Preamble for Markovian memory — 422
15.4.2 Classical capacity — 429

16 Channels with long-term memory — 441
16.1 The model — 441
16.2 Classical product state capacity — 442
16.2.1 Infinite-dimensional classical capacity — 443
16.2.2 Constrained classical capacity in infinite dimensions — 455
16.3 Entanglement-assisted classical capacity — 458
16.3.1 Unconstrained case — 461
16.3.2 Energy constrained case — 466

Bibliography — 471

Index — 479

1 Basic notation and preliminaries

The purposes of this chapter are twofold: (1) to set basic notation that are universal throughout the book and (2) to review relevant preliminaries from functional analysis and operators theory that will be used as tools in the following chapters. It is not intended for readers to dwell too much into this chapter because its details can be found in most of the standard functional analysis textbooks or monographs. The readers who are familiar with the topics can use it as a reminder and quickly glance through them. The readers who are not familiar with the topics and would like to read the proofs omitted in this chapter are encouraged to consult standard functional analysis textbooks or reference books such as Rudin [134], Conway [28], Reed and Simon [128], Yosida [182] and a more recent monograph by van Neerven [171]. To shorten the presentation of this chapter, other mathematical tools will be presented and mingled with topics of quantum information when and where they are needed throughout the book.

– It is recommended that readers skip the proofs of the results in this chapter at the first reading and revisit them when it is needed at a later time.

1.1 Complex Hilbert and Banach spaces

This section serves as a review of complex Hilbert and Banach spaces. Some of the frequently used theorems and/or propositions are stated without a proof.

Complex Hilbert spaces play an important role in the description of quantum systems. As mentioned in Chang [24], every quantum system is associated with an infinite-dimensional separable or a finite-dimensional complex Hilbert space, which consists of the states of the quantum system. In physics terminology, the Hilbert space is usually referred to as the space of (pure) states. Throughout this monograph, the mathematical description of a quantum system shall be based on a certain complex (separable) Hilbert space $\mathbb{H}$ and, therefore, the quantum system will simply be denoted by $\mathbb{H}$.

The quantum system $\mathbb{H}$ is said to be a finite-dimensional system if $\mathbb{H}$ is a finite-dimensional complex Hilbert space. Otherwise, the quantum system $\mathbb{H}$ is said to be an infinite-dimensional system.

We first set some basic notation below.

Let $\mathbb{R}$ and $\mathbb{C}$ denote the field of real numbers and the field of complex numbers, respectively. If $z = x + iy \in \mathbb{C}$, where $x, y \in \mathbb{R}$, let $\bar{z} = x - iy \in \mathbb{C}$ and $|z| = \sqrt{x^2 + y^2} \in \mathbb{R}_+$ denote the complex conjugate and the modulus of the complex number $z \in \mathbb{C}$, respectively. In this case, $x = \mathbb{R}(z)$ is the real part of z and $y = \mathbb{J}(z)$ is the imaginary part of z . Throughout the end, elements in $\mathbb{R}$ or $\mathbb{C}$ shall be denoted by lowercase letters such as a, b or c and sometimes lower case Greek alphabets such as λ and α.

We also use the following conventional notation throughout the book:

https://doi.org/10.1515/9783110788105-001

- $\mathbb{N}$ is the set of all *natural numbers*, positive integers, i. e., $\mathbb{N} = \{1, 2, \cdots, n, \cdots\}$.
- $\mathbb{Z}$ is the set of all integers, i. e., $\mathbb{Z} = \{\cdots, -2, -1, 0, 1, 2, \cdots\}$.
- $\mathbb{Z}_+$ is the set of nonnegative integers, i. e., $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$.
- $\mathbb{R}_+ = \{c \in \mathbb{R} \mid c \geq 0\}$.
- For $-\infty < a < b < +\infty$, we use the usual convention for closed, open and half-open *intervals* on the real line $\mathbb{R}$ such as $[a, b]$, $[a, b[$, $]a, b]$, $]-\infty, a]$, $]-\infty, a[$, $[b, +\infty[$ and $]b, \infty[$, etc.

Let $\mathbb{H}$ be a (generic) Hilbert space over the field of complex numbers $\mathbb{C}$ and be referred to as a complex Hilbert space throughout the end. The complex Hilbert space $\mathbb{H}$ shall be equipped with the (Hermitian) Hilbertian inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}} : \mathbb{H} \times \mathbb{H} \to \mathbb{C}$ that satisfies the following conditions:
(i) (linearity in second argument)

$$\langle \phi, a\varphi + b\varsigma \rangle_{\mathbb{H}} = a\langle \phi, \varphi \rangle_{\mathbb{H}} + b\langle \phi, \varsigma \rangle_{\mathbb{H}}, \quad \forall a, b \in \mathbb{C} \text{ and } \forall \phi, \varphi, \varsigma \in \mathbb{H},$$

(ii) (conjugate-linearity in first argument)

$$\langle a\phi + b\varphi, \zeta \rangle_{\mathbb{H}} = \bar{a}\langle \phi, \zeta \rangle_{\mathbb{H}} + \bar{b}\langle \varphi, \zeta \rangle_{\mathbb{H}}, \quad \forall a, b \in \mathbb{C} \text{ and } \forall \phi, \varphi, \zeta \in \mathbb{H}.$$

(iii) (conjugate symmetry)

$$\overline{\langle \phi, \varphi \rangle_{\mathbb{H}}} = \langle \varphi, \phi \rangle_{\mathbb{H}}, \quad \forall \phi, \varphi \in \mathbb{H}, \text{ and}$$

(iv) (positive definiteness)

$$\langle \phi, \phi \rangle_{\mathbb{H}} \geq 0, \quad \forall \phi \in \mathbb{H}, \text{ and } \langle \phi, \phi \rangle_{\mathbb{H}} = 0 \text{ if and only if } \phi = 0.$$

The Hilbertian norm $\| \cdot \|_{\mathbb{H}} : \mathbb{H} \to \mathbb{R}$ corresponding to the inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ is defined by

$$\|\psi\|_{\mathbb{H}} = \sqrt{\langle \psi, \psi \rangle_{\mathbb{H}}}, \quad \forall \psi \in \mathbb{H}.$$

The complex conjugation $\bar{\phi}$ of $\phi \in \mathbb{H}$ can be considered as any operation on $\mathbb{H}$ that satisfies the following two properties:
1. $\overline{a\phi + b\psi} = \bar{a}\bar{\phi} + \bar{b}\bar{\psi}$, $\forall a, b \in \mathbb{C}$ and $\forall \phi, \phi \in \mathbb{H}$;
2. $\bar{\bar{\phi}} = \phi$, $\forall \phi \in \mathbb{H}$.

A vector $\phi \in \mathbb{H}$ is said to be a real vector if $\bar{\phi} = \phi$. The set of all real vectors forms a real subspace of $\mathbb{H}$, denoted by $\mathfrak{R}(\mathbb{H})$. This is clearly not a complex subspace. In fact, every $\varphi \in \mathbb{H}$ can be written uniquely as $\varphi = \phi + \imath\psi$ with $\phi, \psi \in \mathfrak{R}(\mathbb{H})$, where $\imath = \sqrt{-1}$ is the *imaginary unit*.

While complex Hilbert spaces are the main framework, we will be working within, and we occasionally also work with a more general complete normed linear space over the complex field, called a complex Banach space $\mathbb{B}$ equipped with the (Banach) norm $\|\cdot\|_{\mathbb{B}}$. Recall that the function $\|\cdot\|_{\mathbb{B}} : \mathbb{B} \to \mathbb{R}_+$ is said to be a (Banach) norm if the following conditions are satisfied:

(i) $\|\varphi\|_{\mathbb{B}} \geq 0$ for all $\varphi \in \mathbb{B}$;
(ii) $\|c\varphi\|_{\mathbb{B}} = |c|\|\varphi\|_{\mathbb{B}}$ for all $c \in \mathbb{C}$ and $\varphi \in \mathbb{B}$;
(iii) $\|\phi + \varphi\|_{\mathbb{B}} \leq \|\phi\|_{\mathbb{B}} + \|\varphi\|_{\mathbb{B}}$ for all $\phi, \varphi \in \mathbb{B}$ and
(iv) $\|\varphi\|_{\mathbb{B}} = 0$ if and only if $\varphi = 0$, where 0 is the zero vector in $\mathbb{B}$.

Recall that a linear vector space $\mathbb{X}$ with norm $\|\cdot\|_{\mathbb{X}}$ over a complex field is said to be complete if every Cauchy sequence converges to an element/vector in the space (see, e. g., Rudin [134], Conway [28], Reed and Simon [128] and Yosida [182]). A sequence $(\varphi_n)_{n=1}^{+\infty} \subset \mathbb{X}$ is said to be a Cauchy sequence if for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that $\|\varphi_n - \varphi_m\|_{\mathbb{X}} < \epsilon$ for all $m, n \geq N$.

It is clear that a complex Hilbert space is a special case of a complex Banach space under the Hilbertian norm $\|\cdot\|_{\mathbb{H}}$. However, the converse is not true, because a Banach space lacks Hilbertian inner product structure, in general.

Throughout the end, elements (or vectors) of $\mathbb{X}$ (here $\mathbb{X} = \mathbb{H}$ or $\mathbb{B}$) shall be denoted by lowercase Greek symbols such as ϕ, φ and ζ and, occasionally, the lowercase letters such as u, v and w. Using Dirac's "bra" and "ket" notation (see Dirac [38]), those vectors in the complex Hilbert space $\mathbb{H}$ are also written as $|\phi\rangle_{\mathbb{H}}$, $|\varphi\rangle_{\mathbb{H}}$, $|\zeta\rangle_{\mathbb{H}}$, $|u\rangle_{\mathbb{H}}$, $|v\rangle_{\mathbb{H}}$, $|w\rangle_{\mathbb{H}}$, (or simply $|\phi\rangle$, $|\varphi\rangle$, $|\zeta\rangle$, $|u\rangle$, $|v\rangle$, $|w\rangle$ when there is no danger of ambiguity), etc. These notation are used interchangeably throughout to the end. We also often write $|\phi\rangle_A = |\phi\rangle_{\mathbb{H}_A}$, $\phi\rangle_B = |\phi\rangle_{\mathbb{H}_B}$, etc., for vectors in Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$ with index A and B, etc.

We review some of the topological concepts and properties of $\mathbb{H}$ or $\mathbb{B}$ below.

For $\mathbb{X} = \mathbb{H}$ or $\mathbb{B}$, the closure of $A \subseteq \mathbb{X}$ in the norm $\|\cdot\|_{\mathbb{X}}$ is denoted by $\bar{A}$. A subset $A \subseteq \mathbb{X}$ is said to be dense in $\mathbb{X}$ if $\bar{A} = \mathbb{X}$. In this case, every element in $\mathbb{X}$ can be approximated by elements from A. More precisely, for every $\psi \in \mathbb{X}$ there exists a sequence $(\psi_n)_{n=1}^{+\infty} \subseteq A$ such that $\lim_{n\to+\infty} \|\psi_n - \psi\|_{\mathbb{X}} = 0$. A subset $A \subset \mathbb{X}$ is said to be a compact subset of $\mathbb{X}$ if every open covering of A has a finite subcovering. That is, if $A = \cup_{\lambda\in\Lambda} O_\lambda$, where O_λ, $\lambda \in \Lambda$, is an open subset of $\mathbb{X}$, then there exist finite number of open sets O_{λ_i}, $\lambda_i \in \Lambda$ for $i = 1, 2, \ldots, N$ such that $A = \cup_{i=1}^{N} O_{\lambda_i}$. The above is a formal definition for compactness of a set (see Rudin [134], Conway [28]). However, for convenience, we often adapt the following definition whenever there are sequences of elements in A are involved: A is compact in $\mathbb{X}$ if every sequence $(\varphi_n)_{n=1}^{+\infty}$ in A has a subsequence $(\varphi_{n_k})_{k=1}^{+\infty}$ that converges to some vector φ in A under the norm $\|\cdot\|_{\mathbb{X}}$. Note that $A \subset \mathbb{H}$ is said to be relatively compact in $\mathbb{H}$ if $\bar{A}$ (the closure of A) is a compact subset of $\mathbb{X}$. We say that $A \subset \mathbb{H}$ is weakly compact if every sequence $(\phi_n)_{n=1}^{+\infty}$ in A has a subsequence, which converges weakly to a vector in $\bar{A}$ (see (1.4) for the definition of weak convergence sequence below in a complex Hilbert space).

The complex Hilbert or Banach space $\mathbb{X}$ is said to be separable if it contains a countable dense subset. Examples of separable Hilbert or Banach spaces are abundant and we often assume that the Hilbert spaces that describe the quantum systems of interest are separable. Recall that a subset A of $\mathbb{X}$ is said to be total in $\mathbb{X}$ if $\text{span}(A)$ is dense in $\mathbb{X}$, where $\text{span}(A)$ is the space spanned by A. In other words, $\text{span}(A)$ is the space of all finite linear combinations of elements in A and is also referred to as the linear manifold generated by A.

Two vectors ψ and ϕ in Hilbert space $\mathbb{H}$ are said to be *orthogonal* if $\langle\psi,\phi\rangle_{\mathbb{H}} = 0$. In this case, we denote $\psi \perp \phi$. A set $A \subset \mathbb{H}$ is called an orthogonal set of vectors if $\psi \perp \phi$ for all $\psi,\phi \in A$ and $\psi \neq \phi$. An orthogonal set $A \subset \mathbb{H}$ is an orthonormal set if $\|\psi\|_{\mathbb{H}} = 1$ for all $\psi \in A$. An *orthonormal basis* A for $\mathbb{H}$ is a maximal orthonormal set. That is, if $B \subset \mathbb{H}$ is such that $A \subset B$ then B is not an orthonormal set. A set of mutually orthogonal unit vectors in a (possibly infinite-dimensional) vector space is said to be a *complete orthonormal basis* if it is contained in no larger such set. In other words, no other nonzero vector is orthogonal to all the vectors in the set. A complete orthonormal set is also known as a closed orthonormal set.

If $A \subset \mathbb{H}$, we set

$$A^{\perp} = \{\psi \in \mathbb{H} \mid \langle\psi,\phi\rangle_{\mathbb{H}} = 0, \forall\phi \in A\}.$$

If A and B are subsets of $\mathbb{H}$ such that $A \subseteq B$, then it is easy to verify that $B^{\perp} \subseteq A^{\perp}$ and $A^{\perp\perp} = \bar{A}$ (the closure of A).

Hilbert space $\mathbb{H}$ is said to be N-dimensional if its orthonormal basis A consists of N elements (vectors) for some positive integer N. The separable Hilbert space is said to be infinite-dimensional if its orthonormal basis A consists of infinitely but countably many elements (vectors).

Definition 1.1.1. Let $\mathbb{H}$ and $\mathbb{K}$ be two complex Hilbert spaces with inner products $\langle\cdot,\cdot\rangle_{\mathbb{H}}$ and $\langle\cdot,\cdot\rangle_{\mathbb{K}}$, respectively. The Hilbert spaces $\mathbb{H}$ and $\mathbb{K}$ are said to be isomorphic if there is a bijective (one-to-one and onto) linear mapping $\mathbf{U} : \mathbb{H} \to \mathbb{K}$ such that

$$\langle\mathbf{U}\phi,\mathbf{U}\varphi\rangle_{\mathbb{K}} = \langle\phi,\varphi\rangle_{\mathbb{H}}, \quad \forall\phi,\varphi \in \mathbb{H}.$$

In this case, the linear mapping $\mathbf{U}$ is said to be an isomorphism.

It is clear that every complex Hilbert space $\mathbb{H}$ is isomorphic to itself with the identity operator $\mathbf{I}_{\mathbb{H}} : \mathbb{H} \to \mathbb{H}$ being an isomorphism. Note that $\mathbf{I}_{\mathbb{H}}$ is the identity operator if $\mathbf{I}_{\mathbb{H}}\phi = \phi$ for all $\phi \in \mathbb{H}$.

Some of the widely known and frequently used Hilbert spaces are given below.

Example 1.1. $\mathbb{C}^N$, the space of N-component complex vectors, is an N-dimensional Hilbert space equipped with the inner product $\langle\cdot,\cdot\rangle_{\mathbb{C}^N} : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}$ defined by $\langle a,b\rangle_{\mathbb{C}^N} = \sum_{i=1}^N \overline{a_i} b_i$ for all $a = (a_1,a_2,\dots,a_N)$ and $b = (b_1,b_2,\dots,b_N)$ in $\mathbb{C}^N$, where $\overline{a_i}$ is the complex conjugate of a_i.

Example 1.2. The space of square summable complex sequences,

$$l^2(\mathbb{N};\mathbb{C}) = \left\{(a_n)_{n=1}^{+\infty} | a_n \in \mathbb{C} \text{ for all } n \text{ and } \sum_{n=1}^{+\infty} |a_n|^2 < +\infty\right\},$$

is an infinite-dimensional complex Hilbert space equipped with the inner product

$$\langle\cdot,\cdot\rangle_{l^2} : l^2(\mathbb{N};\mathbb{C}) \times l^2(\mathbb{N};\mathbb{C}) \to \mathbb{C}$$

defined by $\langle (a_n)_{n=1}^{+\infty}, (b_n)_{n=1}^{+\infty}\rangle_{l^2} = \sum_{n=1}^{+\infty} \overline{a_n} b_n$ for all sequences $(a_n)_{n=1}^{+\infty}$ and $(b_n)_{n=1}^{+\infty} \in l^2(\mathbb{N};\mathbb{C})$. From a functional analysis point of view, all infinite-dimensional separable complex Hilbert spaces are isomorphic to $l^2(\mathbb{N};\mathbb{C})$. Nevertheless, finite-dimensional (say dimension $= N$) Hilbert spaces occur in quantum systems as well. In this case, the N-dimensional complex Hilbert space $\mathbb{H}$ can be identified as $\mathbb{C}^N$.

For notational simplicity, we often write $\langle\cdot,\cdot\rangle_{l^2}$ as $\langle\cdot,\cdot\rangle_2$.

Example 1.3. Let $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ be a Borel measurable space, where $\mathbb{X}$ is a metric space, $\mathcal{B}(\mathbb{X})$ is the Borel subsets of $\mathbb{X}$ and μ is a σ-finite measure on $\mathcal{B}(\mathbb{X})$. Let $L^2(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu; \mathbb{C})$ (or simply $L^2(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ for notational simplicity) be the space of complex-valued measurable functions $f : \mathbb{X} \to \mathbb{C}$ such that $\int_{\mathbb{X}} |f(x)|^2 \mu(dx) < \infty$. The space $L^2(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu; \mathbb{C})$ is an infinite-dimensional separable complex Hilbert space equipped with the inner product

$$\langle f, g\rangle_{L^2} = \int_{\mathbb{X}} \overline{f(x)} g(x) \mu(dx), \quad \forall f, g \in L^2(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu).$$

For notational simplicity, we often use the same notation as that for l^2 and write $\langle\cdot,\cdot\rangle_{L^2}$ as $\langle\cdot,\cdot\rangle_2$ when there is no ambiguity that can arise.

The concept of a Borel measurable space and integral of measurable functions will be briefly introduced later when we define positive operator-valued measures in Subsection 2.6.1 and their details can be found in [134] and [28], etc.

All of these Hilbert spaces given in Examples 1.1–1.3 are separable.

If $(e_n)_{n=1}^{+\infty}$ (or $(|e_n\rangle_{\mathbb{H}})_{n=1}^{+\infty}$ in *Dirac's notation*) is a complete orthonormal basis of an infinite-dimensional separable Hilbert space $\mathbb{H}$, then every vector ψ (or $|\psi\rangle_{\mathbb{H}}$ in Dirac's notation) in $\mathbb{H}$ can be expanded in terms of $(e_n)_{n=1}^{+\infty}$ as

$$|\psi\rangle_{\mathbb{H}} := \psi = \sum_{n=1}^{+\infty} \langle\psi, e_n\rangle_{\mathbb{H}} e_n, \quad \forall \psi \in \mathbb{H},$$

and the following Parsaval equation holds:

$$\|\psi\|_{\mathbb{H}}^2 = \sum_{n=1}^{+\infty} \langle\psi, e_n\rangle_{\mathbb{H}}^2. \tag{1.1}$$

The above expansion also applies to finite-dimensional Hilbert spaces in a trivial manner.

The following well-known *Cauchy–Schwarz inequality* in $\mathbb{H}$ will be used frequently as well:

$$\left|\langle \phi, \varphi \rangle_{\mathbb{H}}\right| \leq \|\phi\|_{\mathbb{H}} \|\varphi\|_{\mathbb{H}}, \quad \forall \phi, \varphi \in \mathbb{H}. \tag{1.2}$$

A sequence $(\phi_n)_{n=1}^{+\infty}$ in a complex Hilbert space $\mathbb{H}$ is said to *converge strongly* to $\phi \in \mathbb{H}$ (or equivalent converges in $\|\cdot\|_{\mathbb{H}}$-norm) and denoted by $(s) \lim_{n\to+\infty} \phi_n = \phi$ if

$$\lim_{n\to+\infty} \|\phi_n - \phi\|_{\mathbb{H}} = 0, \tag{1.3}$$

and the sequence $(\phi_n)_{n=1}^{+\infty}$ is said to *converge weakly* to $\phi \in \mathbb{H}$ and denoted by $(w) \lim_{n\to+\infty} \phi_n = \phi$ if

$$\lim_{n\to+\infty} \langle \phi_n, \psi \rangle_{\mathbb{H}} = \langle \phi, \psi \rangle_{\mathbb{H}} \quad \text{or} \quad \lim_{n\to+\infty} \langle \psi, \phi_n \rangle_{\mathbb{H}} = \langle \psi, \phi \rangle_{\mathbb{H}}, \quad \forall \psi \in \mathbb{H}. \tag{1.4}$$

It is clear that strong convergence implies weak convergence. That is,

$$(s) \lim_{n\to+\infty} \phi_n = \phi \Rightarrow (w) \lim_{n\to+\infty} \phi_n = \phi, \tag{1.5}$$

or precisely,

$$\lim_{n\to\infty} \|\phi_n - \phi\|_{\mathbb{H}} = 0 \Rightarrow \lim_{n\to\infty} \langle \phi_n - \phi, \psi \rangle_{\mathbb{H}} = 0, \quad \forall \psi \in \mathbb{H}. \tag{1.6}$$

This is because

$$\begin{aligned}\lim_{n\to\infty} \left|\langle \phi_n, \psi \rangle_{\mathbb{H}} - \langle \phi, \psi \rangle_{\mathbb{H}}\right| &= \lim_{n\to\infty} \left|\langle \phi_n - \phi, \psi \rangle_{\mathbb{H}}\right| \\ &\leq \|\psi\|_{\mathbb{H}} \lim_{n\to\infty} \|\phi_n - \phi\|_{\mathbb{H}} = 0,\end{aligned}$$

where the Cauchy–Schwarz inequality (1.2) is applied to the last inequality in the above.

Next, we state without proof that a weakly convergent sequence in a complex Hilbert space $\mathbb{H}$ is bounded, and gives a useful necessary and sufficient condition for weak convergence. We recall equation (1.4) for the definition of a weak converge sequence. The proof of the following two results can be found in Rudin [134].

Theorem 1.1.2. *Suppose that $(\phi_n)_{n=1}^{+\infty}$ is a sequence of vectors in a Hilbert space $\mathbb{H}$ and $\mathbb{D}$ is a dense subset of $\mathbb{H}$. Then $(\phi_n)_{n=1}^{+\infty}$ converges strongly to ϕ if and only if:*
1. *If there exists a constant $M > 0$ such that $\|\phi_n\|_{\mathbb{H}} \leq M$ for all $n \geq 1$ and*
2. *$\langle \phi_n, \psi \rangle_{\mathbb{H}} \to \langle \phi, \psi \rangle_{\mathbb{H}}$ as $n \to +\infty$ for all $\psi \in \mathbb{D}$.*

Proposition 1.1.3. *If the sequence* $(\phi_n)_{n=1}^{+\infty} \subset \mathbb{H}$ *converges weakly to* $\phi \in \mathbb{H}$*, then*

$$\|\phi\|_{\mathbb{H}} \leq \liminf_{n\to+\infty} \|\phi_n\|_{\mathbb{H}}. \tag{1.7}$$

If, in addition,

$$\lim_{n\to+\infty} \|\phi_n\|_{\mathbb{H}} = \|\phi\|_{\mathbb{H}}, \tag{1.8}$$

then the sequence $(\phi_n)_{n=1}^{+\infty}$ *converges to* ϕ *strongly.*

A proof of the theorem stated below can be found in any standard functional analysis textbook (see, e. g., Section 3.15, p. 18 of Rudin [134] and Chapter 5, Section 3 of Conway [28]) and is therefore omitted here.

Theorem 1.1.4 (Banach–Alaoglu theorem). *The closed unit ball of a Hilbert space is weakly compact.*

Throughout to the end of this book, we assume that all complex Hilbert spaces are separable (and hence a countable complete orthonormal basis exists for the space) without specifically mentioned.

1.2 Linear operators and their adjoints

Let $\mathbb{X}$ and $\mathbb{Y}$ be two separable complex Banach spaces equipped with Banach norms $\|\cdot\|_{\mathbb{X}}$ and $\|\cdot\|_{\mathbb{Y}}$, respectively.

A map (or *transformation*) $\mathbf{T} : \mathrm{dom}(\mathbf{T}) \subseteq \mathbb{X} \to \mathbb{Y}$ is said to be *linear* if

$$a\phi + b\psi \in \mathrm{dom}(\mathbf{T})$$

and

$$\mathbf{T}(a\phi + b\psi) = a\mathbf{T}(\phi) + b\mathbf{T}(\psi), \quad \forall a, b \in \mathbb{C} \text{ and } \forall \phi, \psi \in \mathrm{dom}(\mathbf{T}),$$

where $\mathrm{dom}(\mathbf{T}) \subseteq \mathbb{X}$ is called the *domain* of $\mathbf{T}$. Note that $\mathrm{dom}(\mathbf{T}) \neq \mathbb{X}$, in general. However, it is known that $\mathrm{dom}(\mathbf{T})$ is a dense subset of $\mathbb{X}$.

The following are some known cases where $\mathrm{dom}(\mathbf{T}) = \mathbb{X}$:

(i) $\mathbf{T}$ is a bounded linear operator (see the following subsection for the definition of a bounded linear operator);

(ii) $\mathbb{X} = \mathbb{H}$ is a Hilbert space and $\dim(\mathbb{H}) < +\infty$, where $\dim(\mathbb{H})$ denotes the dimension of $\mathbb{H}$. In this case, every linear operator is a bounded linear operator (see (1.10) below for the definition of a bounded linear operator).

Assuming $\mathrm{dom}(\mathbf{T})$ is dense in $\mathbb{X}$ or $\mathrm{dom}(\mathbf{T}) = \mathbb{X}$, the collection of such linear maps will be denoted by $\mathfrak{L}(\mathbb{X}, \mathbb{Y})$. The linear map $\mathbf{T}$ will be called a linear operator if $\mathrm{dom}(\mathbf{T}) \subseteq \mathbb{X} = \mathbb{Y}$. The collection of linear operators will be denoted by $\mathfrak{L}(\mathbb{X})$.

Throughout to the end, linear maps/operators on a complex Banach or Hilbert space will be denoted by boldfaced letters such as $\mathbf{S}, \mathbf{T}, \mathbf{a}, \mathbf{b}, \mathbf{X}, \mathbf{Y}$, etc., and a linear map $\mathbf{T} \in \mathfrak{L}(\mathbb{X}, \mathbb{Y})$ acting on a vector $\phi \in \mathbb{X}$ will be denoted by either $\mathbf{T}\phi$ or $\mathbf{T}(\phi)$.

Let $\mathbf{T} : \mathrm{dom}(\mathbf{T}) \to \mathrm{range}(\mathbf{T})$ (where $\mathrm{range}(\mathbf{T}) := \{\mathbf{T}\phi \in \mathbb{Y} \mid \phi \in \mathrm{dom}(\mathbf{T}) \subset \mathbb{X}\}$ denotes the range of $\mathbf{T}$) be a bijective (one-to-one and onto) linear map. A linear map $\mathbf{S} : \mathrm{range}(\mathbf{T}) \to \mathrm{dom}(\mathbf{S})$ is said to be the *inverse* of $\mathbf{T}$ if $\mathbf{S} \circ \mathbf{T} = \mathbf{I}_{\mathbb{X}}$ and $\mathbf{T} \circ \mathbf{S} = \mathbf{I}_{\mathbb{Y}}$, where $\mathbf{I}_{\mathbb{X}}$ and $\mathbb{I}_{\mathbb{Y}}$ are the identity operator on $\mathbb{X}$ and $\mathbb{Y}$, respectively. That is, $\mathbf{I}_{\mathbb{X}}(x) = x$ for all $x \in \mathbb{X}$ and $\mathbf{I}_{\mathbb{Y}}(y) = y$ for all $y \in \mathbb{Y}$. In this case, we write $\mathbf{S} = \mathbf{T}^{-1}$.

For a linear map $\mathbf{T} : \mathrm{dom}(\mathbf{T}) \subset \mathbb{H} \to \mathbb{K}$, where $\mathbb{H}$ and $\mathbb{K}$ are complex Hilbert spaces, $\ker(\mathbf{T})$ (the kernel of $\mathbf{T}$), $\mathrm{range}(\mathbf{T})$ (the range of $\mathbf{T}$) and $\mathrm{supp}(\mathbf{T})$ (the support of $\mathbf{T}$) are defined as:

$$\begin{aligned}
\ker(\mathbf{T}) &= \{\phi \in \mathrm{dom}(\mathbf{T}) \subset \mathbb{H} \mid \mathbf{T}\phi = 0\};\\
\mathrm{range}(\mathbf{T}) &= \{\psi \in \mathbb{K} \mid \psi = \mathbf{T}\phi \text{ for some } \phi \in \mathrm{dom}(\mathbf{T}) \subset \mathbb{H}\};\\
\mathrm{supp}(\mathbf{T}) &= (\ker(\mathbf{T}))^{\perp} := \{\psi \in \mathbb{H} \mid \langle \psi, \phi\rangle_{\mathbb{H}} = 0, \forall \phi \in \ker(\mathbf{T})\}.
\end{aligned}$$

A linear map $\mathbf{T} : \mathrm{dom}(\mathbf{T}) \to \mathbb{K}$ is said to be *closed* if its graph,

$$\mathrm{graph}(\mathbf{T}) := \{(\phi, \mathbf{T}(\phi)) \in \mathbb{H} \times \mathbb{K} \mid \phi \in \mathrm{dom}(\mathbf{T})\},$$

is a closed subset of $\mathbb{H} \times \mathbb{K}$.

1.2.1 Adjoint and self-adjoint operators

Although one can consider adjointness of a linear map from one Banach space to another, we consider for simplicity in the following adjoint map of the linear map $\mathbf{T}$, when $\mathbb{H}$ and $\mathbb{K}$ are complex Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and $\langle \cdot, \cdot \rangle_{\mathbb{K}}$, respectively.

If $\mathrm{dom}(\mathbf{T})$, the domain of $\mathbf{T}$, is dense in $\mathbb{H}$, i. e., $\overline{\mathrm{dom}(\mathbf{T})} = \mathbb{H}$. Let $\mathrm{dom}(\mathbf{T}^*)$ be the subset of $\mathbb{K}$ defined by

$$\mathrm{dom}(\mathbf{T}^*) = \{\phi \in \mathbb{K} \mid \exists \varphi \in \mathbb{H} \text{ such that } \langle \mathbf{T}\zeta, \phi\rangle_{\mathbb{K}} = \langle \zeta, \varphi\rangle_{\mathbb{H}}, \forall \zeta \in \mathrm{dom}(\mathbf{T})\}.$$

For each such $\phi \in \mathrm{dom}(\mathbf{T}^*)$, we define $\mathbf{T}^*\phi = \varphi$. We call the operator $\mathbf{T}^* : \mathrm{dom}(\mathbf{T}^*) \subset \mathbb{K} \to \mathbb{H}$ the adjoint of $\mathbf{T}$. In this case,

$$\langle \mathbf{T}\zeta, \phi\rangle_{\mathbb{K}} = \langle \zeta, \mathbf{T}^*\phi\rangle_{\mathbb{H}} \quad \forall \zeta \in \mathrm{dom}(\mathbf{T}) \text{ and } \phi \in \mathrm{dom}(\mathbf{T}^*). \tag{1.9}$$

If $\mathbf{S}, \mathbf{T} \in \mathfrak{L}(\mathbb{H})$ (i. e., $\mathbf{S}$ and $\mathbf{T}$ are linear operators on Hilbert space $\mathbb{H}$), then:
(i) $\mathbf{T}^{**} := (\mathbf{T}^*)^* = \mathbf{T}$,
(ii) $(c\mathbf{T} + \mathbf{S})^* = \bar{c}\mathbf{T}^* + \mathbf{S}^*$ and
(iii) $(\mathbf{ST})^* = \mathbf{T}^*\mathbf{S}^*$, where $(\mathbf{ST})\phi := (\mathbf{S} \circ \mathbf{T})\phi = \mathbf{S}(\mathbf{T}\phi), \forall \phi \in \mathbb{H}$.

The linear operator $\mathbf{T} \in \mathfrak{L}(\mathbb{H})$ is said to be *symmetric* if $\mathrm{dom}(\mathbf{T}) \subset \mathrm{dom}(\mathbf{T}^*)$ and $\mathbf{T}\zeta = \mathbf{T}^*\zeta$ for all $\zeta \in \mathrm{dom}(\mathbf{T})$. Equivalently, $\mathbf{T}$ is symmetric if and only if $\langle \mathbf{T}\zeta, \varphi\rangle_{\mathbb{H}} = \langle \zeta, \mathbf{T}\varphi\rangle_{\mathbb{H}}$ for all $\zeta, \varphi \in \mathrm{dom}(\mathbf{T})$. The linear operator $\mathbf{T} \in \mathfrak{L}(\mathbb{H})$ is called *self-adjoint* if $\mathbf{T} = \mathbf{T}^*$, i. e., $\mathbf{T}$ is symmetric and $\mathrm{dom}(\mathbf{T}) = \mathrm{dom}(\mathbf{T}^*)$.

The collection of linear self-adjoint operators will be denoted by $\mathfrak{L}_{\mathrm{sa}}(\mathbb{H})$.

Example 1.4. We consider the N-dimensional complex Hilbert space $\mathbb{C}^N$. Every linear operator $\mathbf{T} : \mathbb{C}^N \to \mathbb{C}^N$ can be represented as an $N \times N$ complex matrix denoted by

$$\mathbf{T} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix}.$$

(i) $\mathbf{T}^*$, the adjoint of $\mathbf{T}$, can be expressed as

$$\mathbf{T}^* = \begin{bmatrix} \overline{a_{11}} & \overline{a_{12}} & \dots & \overline{a_{1N}} \\ \overline{a_{21}} & \overline{a_{22}} & \dots & \overline{a_{2N}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{N1}} & \overline{a_{N2}} & \dots & \overline{a_{NN}} \end{bmatrix}^\top = \begin{bmatrix} \overline{a_{11}} & \overline{a_{21}} & \dots & \overline{a_{N1}} \\ \overline{a_{12}} & \overline{a_{22}} & \dots & \overline{a_{N2}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1N}} & \overline{a_{2N}} & \dots & \overline{a_{NN}} \end{bmatrix},$$

where $\overline{a_{ij}}$ is the complex conjugate of a_{ij} and $A^\top$ denotes the transpose of the matrix A.
(ii) $\mathbf{T}$ is said to be a *Hermitian* if it is self-adjoint, i. e., $\mathbf{T} = \mathbf{T}^*$.

1.2.2 Bounded linear operators

Let $\mathbb{X}$ and $\mathbb{Y}$ be complex Banach spaces (or Hilbert spaces) equipped with the norm $\|\cdot\|_{\mathbb{X}}$ and $\|\cdot\|_{\mathbb{Y}}$, respectively.

A linear map $\mathbf{T} : \mathrm{dom}(\mathbf{T}) \subset \mathbb{X} \to \mathbb{Y}$ is said to be a bounded linear map if $\mathbf{T}(B)$ is a bounded subset of $\mathbb{Y}$ for every bounded $B \subset \mathrm{dom}(\mathbf{T}) \subset \mathbb{X}$. Equivalently, $\mathbf{T}$ is a bounded linear map if there exists a constant $K > 0$ (that depends on $\mathbf{T}$ only) such that

$$\|\mathbf{T}\psi\|_{\mathbb{Y}} \le K\|\psi\|_{\mathbb{X}}, \quad \forall \psi \in \mathrm{dom}(\mathbf{T}) \subset \mathbb{X}. \tag{1.10}$$

In this case, it can be proved (see, e. g., Reed and Simon [128] and Rudin [134]) that $\mathrm{dom}(\mathbf{T}) = \mathbb{X}$. The collection of bounded linear maps from $\mathbb{X}$ to $\mathbb{Y}$ will be denoted by $\mathfrak{B}(\mathbb{X}, \mathbb{Y})$. If $\mathbb{X} = \mathbb{Y}$, then $\mathfrak{B}(\mathbb{X}, \mathbb{Y})$ can be written as $\mathfrak{B}(\mathbb{X})$ and $\mathbf{T} \in \mathfrak{B}(\mathbb{X})$ will be called a bounded linear operator.

The following is a Banach–Schauder theorem, which is also known as an open mapping theorem. A proof can be found in standard functional analysis texts such as Rudin [133] and is omitted here.

Theorem 1.2.1 (Banach–Schauder theorem). *If $\mathbb{X}$ and $\mathbb{Y}$ are complex Banach spaces or complex Hilbert spaces and $\mathbf{T} : \mathbb{X} \to \mathbb{Y}$ is a surjective (or onto) bounded linear operator, then $\mathbf{T}$ is an open map (i. e., if U is an open set in $\mathbb{X}$, then $\mathbf{T}(U)$ is open in $\mathbb{Y}$).*

We have the following corollary.

Corollary 1.2.2. *If $\mathbb{X}$ and $\mathbb{Y}$ are complex Banach or Hilbert spaces and $\mathbf{T} \in \mathfrak{L}(\mathbb{X}, \mathbb{Y})$ is invertible (i. e., a bijective linear map), then the inverse map, $\mathbf{T}^{-1}$, is bounded, i. e., $\mathbf{T}^{-1} \in \mathfrak{B}(\mathbb{Y}, \mathbb{X})$. (Note that $\mathbf{T}^{-1}$ is automatically linear.)*

If $\mathbf{T} \in \mathfrak{B}(\mathbb{X})$, we define the operator norm $\|\mathbf{T}\|_\infty$ of $\mathbf{T}$ as

$$\|\mathbf{T}\|_\infty = \sup_{\phi \neq 0} \frac{\|\mathbf{T}\phi\|_\mathbb{X}}{\|\phi\|_\mathbb{X}} = \sup_{\|\phi\|_\mathbb{X}=1} \|\mathbf{T}\phi\|_\mathbb{X}. \tag{1.11}$$

There are two trivial bounded linear operators $\mathbf{O}_\mathbb{X}$ (the *zero operator*) and $\mathbf{I}_\mathbb{X}$ (the *identity operator*) that will appear often throughout the book. The *zero operator* $\mathbf{O}_\mathbb{X}$ is the operator that maps every vector $\phi \in \mathbb{X}$ to the zero vector $\mathbf{0}$ in $\mathbb{X}$ (i. e., $\mathbf{O}_\mathbb{X}\phi = \mathbf{0}$ for all $\phi \in \mathbb{X}$) and the identity operator $\mathbf{I}_\mathbb{H}$ is the operator that maps every vector $\phi \in \mathbb{X}$ to itself (i. e., $\mathbf{I}_\mathbb{X}\phi = \phi$ for all $\phi \in \mathbb{X}$.)

A bounded linear operator $\mathbf{T}$ on $\mathbb{H}$ is said to be a contraction if there exists a positive constant $c \leq 1$ such that

$$\|\mathbf{A}\phi\|_\mathbb{H} \leq c\|\phi\|_\mathbb{H}, \quad \forall \phi \in \mathbb{H}. \tag{1.12}$$

In this case, $\|\mathbf{A}\|_\infty \leq 1$.

Recall that if $\mathbb{H}$ is finite-dimensional then all linear operators on $\mathbb{H}$ are bounded linear operators. However, if $\mathbb{H}$ is infinite-dimensional, then there are linear operators that are not necessarily bounded. As noted earlier that if the linear operator $\mathbf{T}$ is unbounded, then $\mathbf{T}$ is normally not defined on the entire Hilbert space $\mathbb{H}$ and we denote its domain by $\mathrm{dom}(\mathbf{T})$ as indicated in the previous section. In this case, $\mathrm{dom}(\mathbf{T})$ is normally a dense subset of $\mathbb{H}$.

The following polarization formula (1.13) can be easily obtained by expanding the right-hand side of the equation.

Lemma 1.2.3 (Polarization formula). *Let $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$ be a bounded linear operator on a complex Hilbert space $\mathbb{H}$. Then the following polarization formula holds:*

$$\begin{aligned}\langle \mathbf{A}\phi, \varphi\rangle_{\mathbb{H}} = \frac{1}{4}\{&\langle \mathbf{A}(\phi+\varphi), (\phi+\varphi)\rangle_{\mathbb{H}} - \langle \mathbf{A}(\phi-\varphi), (\phi-\varphi)\rangle_{\mathbb{H}} \\ &- \iota\langle \mathbf{A}(\phi+\iota\varphi), \phi+\iota\varphi\rangle_{\mathbb{H}} + \iota\langle \mathbf{A}(\phi-\iota\varphi), \phi-\iota\varphi\rangle_{\mathbb{H}}\}\end{aligned} \tag{1.13}$$

for all $\phi, \varphi \in \mathbb{H}$, *where* $\iota = \sqrt{-1}$ *denotes the imaginary unit.*

Proposition 1.2.4. *Let* $\mathbf{A} \in \mathfrak{L}(\mathbb{H})$ *be a linear operator on a Hilbert space* $\mathbb{H}$ *that satisfies* $\langle \mathbf{A}\phi, \psi\rangle_{\mathbb{H}} = \langle \phi, \mathbf{A}\psi\rangle_{\mathbb{H}}$ *for all* $\phi, \psi \in \mathbb{H}$. *Then* $\mathbf{A}$ *is bounded.*

Proof. Suppose to the contrary that $\mathbf{A}$ is unbounded. Then there is a sequence $(\psi_n)_{n=1}^{+\infty}$ in $\mathbb{H}$ such that $\|\psi_n\|_{\mathbb{H}} = 1$ and yet $\|\mathbf{A}\psi_n\|_{\mathbb{H}} \to +\infty$. Consider the sequence of functionals $(f_n)_{n=1}^{+\infty}$ defined by $f_n(\phi) = \langle \mathbf{A}\phi, \psi_n\rangle_{\mathbb{H}}$. For each $\phi \in \mathbb{H}$, $f_n(\phi)$ is bounded for each n, since

$$|f_n(\phi)| = |\langle \mathbf{A}\phi, \psi_n\rangle_{\mathbb{H}}| \le \|\mathbf{A}\phi\|_{\mathbb{H}}\|\psi_n\|_{\mathbb{H}} = \|\mathbf{A}\phi\|_{\mathbb{H}}, \quad \forall n.$$

By the uniform boundedness theorem 2.3.2, the sequence $(\|f_n\|_\infty)_{n=1}^{+\infty}$ is bounded. That is, there exists a $c > 0$ such that $\|f_n\|_\infty < c$ for all n. Finally, note that $\|\mathbf{A}\psi_n\|_{\mathbb{H}}^2 = \langle \mathbf{A}\psi_n, \mathbf{A}\psi_n\rangle_{\mathbb{H}} = |f_n(\mathbf{A}\psi_n)| \le c\|\mathbf{A}\psi_n\|_{\mathbb{H}}$ for all n, and thus $\|\mathbf{A}\psi_n\|_{\mathbb{H}} \le c$, a contradiction to the assumption that $\|\mathbf{A}\psi_n\|_{\mathbb{H}} \to +\infty$. Therefore, $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$. This proves the proposition. □

Lemma 1.2.5. *If* $\mathbf{A} : \mathbb{H} \to \mathbb{H}$ *is a bounded linear operator on* $\mathbb{H}$, *then* $(\ker(\mathbf{A}^*))^\perp$ *is a closed subspace of* $\mathbb{H}$.

Proof. Let $(\phi_n)_{n=1}^{+\infty} \subset (\ker(\mathbf{A}^*))^\perp$ be a sequence that converges strongly to $\phi \in \mathbb{H}$ (i. e., $\lim_{n\to+\infty} \|\phi_n - \phi\|_{\mathbb{H}} = 0$). We want to show that $\phi \in (\ker(\mathbf{A}^*))^\perp$. For any $\psi \in \ker(\mathbf{A}^*)$, we have for all n,

$$0 = \langle \phi_n, \mathbf{A}^*\psi\rangle_{\mathbb{H}} = \langle \mathbf{A}\phi_n, \psi\rangle_{\mathbb{H}}.$$

This shows that

$$\langle \phi, \mathbf{A}^*\psi\rangle_{\mathbb{H}} = \langle \mathbf{A}\phi, \psi\rangle_{\mathbb{H}} = \lim_{n\to+\infty} \langle \mathbf{A}\phi_n, \psi\rangle_{\mathbb{H}} = 0,$$

since $\mathbf{A}$ is a bounded linear operator and $(\phi_n)_{n=1}^{+\infty}$ is strongly convergent (which implies weak convergence by (1.6)). Consequently, $\phi \in (\ker(\mathbf{A}^*))^\perp$. Therefore, $(\ker(\mathbf{A}^*))^\perp$ is a closed subset of $\mathbb{H}$. It is easy to show that by linearity that $\phi, \tilde{\phi} \in (\ker(\mathbf{A}^*)^\perp$ and $a, \tilde{a} \in \mathbb{C}$ imply that $a\phi + \tilde{a}\tilde{\phi} \in (\ker(\mathbf{A}^*)^\perp$. Therefore, $(\ker(\mathbf{A}^*)^\perp$ is a closed subspace. □

The following theorem can be found in Rudin [134], Yoshida [182], van Neerven [171], etc.

Theorem 1.2.6. *If* $\mathbf{A} : \mathbb{H} \to \mathbb{H}$ *is a bounded linear operator, then*

$$\overline{\text{range}(\mathbf{A})} = (\ker(\mathbf{A}^*))^\perp \quad \textit{and} \quad \ker(\mathbf{A}) = (\text{range}(\mathbf{A}))^\perp. \tag{1.14}$$

Proof. 1. We first want to prove that $\overline{\mathrm{range}(\mathbf{A})} \subset (\ker(\mathbf{A}^*))^\perp$. If $\psi \in \mathrm{range}(\mathbf{A})$, then there is a $\phi \in \mathbb{H}$ such that $\psi = \mathbf{A}\phi$. For any $\varphi \in \ker(\mathbf{A}^*)$ (i. e., $\mathbf{A}^*\varphi = \mathbf{0}$), we then have

$$\langle \psi, \varphi \rangle_{\mathbb{H}} = \langle \mathbf{A}\phi, \varphi \rangle_{\mathbb{H}} = \langle \phi, \mathbf{A}^*\varphi \rangle_{\mathbb{H}} = 0.$$

This proves that $\psi \in (\ker(\mathbf{A}^*))^\perp$. Consequently, $\mathrm{range}(\mathbf{A}) \subset (\ker(\mathbf{A}^*))^\perp$. Since $(\ker(\mathbf{A}^*))^\perp$ is closed by Lemma 1.2.5, it follows that $\overline{\mathrm{range}(\mathbf{A})} \subset (\ker(\mathbf{A}^*))^\perp$. On the other hand, if $\psi \in (\mathrm{range}(\mathbf{A}))^\perp$, then for all $\phi \in \mathbb{H}$ we have $0 = \langle \mathbf{A}\phi, \psi \rangle_{\mathbb{H}} = \langle \phi, \mathbf{A}^*\psi \rangle_{\mathbb{H}}$. Therefore, $\mathbf{A}^*\psi = 0$, and hence $\psi \in \ker(\mathbf{A}^*)$. This means that $(\mathrm{range}(\mathbf{A}))^\perp \subset \ker(\mathbf{A}^*)$. By taking the orthogonal complement of this relation, we get

$$(\ker(\mathbf{A}^*))^\perp \subset (\mathrm{range}(\mathbf{A}))^{\perp\perp} = \overline{\mathrm{range}(\mathbf{A})},$$

which proves the first part of (1.14).

2. To prove the second part, we apply the first part to $\mathbf{A}^*$, instead of $\mathbf{A}$, use $(\mathbf{A}^*)^* = \mathbf{A}^{**} = \mathbf{A}$ and take orthogonal complements. This proves the theorem. □

An equivalent formulation of this theorem is that if $\mathbf{A}$ is a bounded linear operator on $\mathbb{H}$, then $\mathbb{H}$ has the orthogonal direct sum $\mathbb{H} = \overline{\mathrm{range}(\mathbf{A})} \oplus \ker(\mathbf{A}^*)$ (see Theorem 1.6.2 for the definition of direct sum).

Lemma 1.2.7. *If* $\mathbf{A}$ *is a bounded self-adjoint operator on a Hilbert space* $\mathbb{H}$, *then*

$$\|\mathbf{A}\|_\infty = \sup_{\|\phi\|_{\mathbb{H}}=1} |\langle \mathbf{A}\phi, \phi \rangle_{\mathbb{H}}| = \sup_{\|\phi\|_{\mathbb{H}}=1} |\langle \phi, \mathbf{A}\phi \rangle_{\mathbb{H}}|.$$

Proof. Let $\|\phi\|_{\mathbb{H}} = 1$ and

$$\alpha = \sup_{\|\phi\|_{\mathbb{H}}=1} |\langle \mathbf{A}\phi, \phi \rangle_{\mathbb{H}}| = \sup_{\|\phi\|_{\mathbb{H}}=1} |\langle \phi, \mathbf{A}\phi \rangle_{\mathbb{H}}|.$$

The inequality $\alpha \le \|\mathbf{A}\|_\infty$ is immediate, since

$$|\langle \mathbf{A}\phi, \phi \rangle_{\mathbb{H}}| \le \|\mathbf{A}\phi\|_{\mathbb{H}} \|\phi\|_{\mathbb{H}} \le \|\mathbf{A}\|_\infty \|\phi\|_{\mathbb{H}}^2,$$

where the first inequality above is by the Cauchy–Schwarz inequality (1.2), and the second inequality is by the fact that $\|\mathbf{A}\phi\|_{\mathbb{H}} \le \|\mathbf{A}\|_\infty \|\phi\|_{\mathbb{H}}$. To prove the reverse inequality, we use the definition of the operator norm $\|\cdot\|_\infty$,

$$\|\mathbf{A}\|_\infty = \sup_{\|\phi\|_{\mathbb{H}}=1} \|\mathbf{A}\phi\|_{\mathbb{H}}.$$

For any $\varphi \in \mathbb{H}$, we have

$$\|\varphi\|_{\mathbb{H}} = \sup_{\|\psi\|_{\mathbb{H}}=1} |\langle \psi, \varphi \rangle_{\mathbb{H}}|.$$

It follows that

$$\|\mathbf{A}\|_\infty = \sup\{|\langle \mathbf{A}\phi, \psi\rangle_{\mathbb{H}}| \mid \|\phi\|_{\mathbb{H}} = 1, \|\psi\|_{\mathbb{H}} = 1\}. \tag{1.15}$$

The polarization formula (see (1.13)) implies that

$$\begin{aligned}\langle \mathbf{A}\phi, \varphi\rangle_{\mathbb{H}} = \frac{1}{4}\{&\mathbf{A}(\phi+\varphi), (\phi+\varphi)\rangle_{\mathbb{H}} - \langle \mathbf{A}(\phi-\varphi), (\phi-\varphi)\rangle_{\mathbb{H}} \\ &- \imath\langle \mathbf{A}(\phi+\imath\varphi), \phi+\imath\varphi\rangle_{\mathbb{H}} + \imath\langle \mathbf{A}(\phi-\imath\varphi), \phi-\imath\varphi\rangle_{\mathbb{H}}\}.\end{aligned}$$

Since $\mathbf{A}$ is self-adjoint, the first two terms are real, and the last two are imaginary. We replace φ by $e^{\imath\theta}\varphi$, where $\theta \in \mathbb{R}$ is chosen so that $\langle e^{\imath\theta}\varphi, \mathbf{A}\phi\rangle_{\mathbb{H}}$ is real. Then the imaginary terms vanish, and we find that

$$\begin{aligned}|\langle \mathbf{A}\phi, \varphi\rangle_{\mathbb{H}}| &= |\langle \phi, \mathbf{A}\varphi\rangle_{\mathbb{H}}| \\ &= \frac{1}{4}|\langle(\phi+\varphi), \mathbf{A}(\phi+\varphi)\rangle_{\mathbb{H}} - \langle \phi-\varphi, \mathbf{A}(\phi-\varphi)\rangle_{\mathbb{H}}|^2 \\ &\le \frac{1}{4}\alpha^2(\|\phi+\varphi\|_{\mathbb{H}}^2 + \|\phi-\varphi\|_{\mathbb{H}}^2)^2 \\ &= \frac{1}{4}\alpha^2(\langle \phi+\varphi, \phi+\varphi\rangle_{\mathbb{H}} + \langle \phi-\varphi, \phi-\varphi\rangle_{\mathbb{H}}) \\ &= \frac{1}{4}\alpha^2(\langle \phi, \phi\rangle_{\mathbb{H}} + \langle \varphi, \varphi\rangle_{\mathbb{H}})^2 \\ &= \frac{1}{4}\alpha^2(\|\phi\|_{\mathbb{H}}^2 + \|\varphi\|_{\mathbb{H}}^2)^2 = \frac{1}{4}\alpha^2(1+1)^2 = \alpha^2,\end{aligned}$$

where we have used the definition of α and the parallelogram law. Using this result and (1.15), we conclude that $\|\mathbf{A}\|_\infty \le \alpha$. This proves the lemma. □

We recall the following closed graph theorem (see, e. g., Rudin [134]) without proof. This theorem will be useful later on.

Theorem 1.2.8 (Closed graph theorem). *Let $\mathbb{X}$ be a complex Banach space. If $\mathbf{T} : \mathbb{X} \to \mathbb{X}$ is a bounded linear operator on $\mathbb{X}$, then the following two statements are equivalent:*

1. *If the sequence $(x_n)_{n=1}^{+\infty} \subset \mathbb{X}$ converges in $\|\cdot\|_{\mathbb{X}}$-norm to some element $x \in \mathbb{X}$, then the sequence $(\mathbf{T}(x_n))_{n=1}^{+\infty}$ converges to $\mathbf{T}(x) \in \mathbb{X}$ in $\|\cdot\|_{\mathbb{X}}$-norm.*
2. *For every sequence $(x_n)_{n=1}^{+\infty}$ in $\mathbb{X}$, if the sequence $(x_n)_{n=1}^{+\infty}$ converges in $\|\cdot\|_{\mathbb{X}}$-norm to some element $x \in \mathbb{X}$ and the sequence $(\mathbf{T}(x_n))_{n=1}^{+\infty}$ converges to some element $y \in \mathbb{X}$ in $\|\cdot\|_{\mathbb{X}}$-norm, then $y = \mathbf{T}(x)$.*

1.3 Positive operators

A linear (but not necessary bounded) operator $\mathbf{T} \in \mathfrak{L}(\mathbb{H})$ on a complex Hilbert space $\mathbb{H}$ is said to be *positive* and to be denoted by $\mathbf{T} \ge \mathbf{0}$ if $\langle \mathbf{T}\varphi, \varphi\rangle_{\mathbb{H}} \ge 0$ for all $\varphi \in \mathrm{dom}(\mathbf{T})$.

Let $\mathbf{T}_1, \mathbf{T}_2 \in \mathfrak{L}(\mathbb{H})$. We say that $\mathbf{T}_1 \geq \mathbf{T}_2$ if $\mathbf{T}_1 - \mathbf{T}_2 := \mathbf{T}_1 + (-\mathbf{T}_2) \geq \mathbf{0}$, where $-\mathbf{T}$ is the linear operator such that $\mathbf{T} + (-\mathbf{T}) = \mathbf{0}$.

The set of all positive linear operators (resp., positive bounded linear operators) on $\mathbb{H}$ will be denoted by $\mathfrak{L}_+(\mathbb{H})$ (resp., $\mathfrak{B}_+(\mathbb{H})$). Both $\mathfrak{L}_+(\mathbb{H})$ and $\mathfrak{B}_+(\mathbb{H})$ are positive cones in the sense that $\mathbf{T} \in \mathfrak{L}_+(\mathbb{H})$ (resp., $\mathfrak{B}_+(\mathbb{H})$) and $c > 0$ imply that $c\mathbf{T} \in \mathfrak{L}_+(\mathbb{H})$ (resp., $\mathfrak{B}_+(\mathbb{H})$).

A sequence of bounded linear operators $(\mathbf{A}_n)_{n=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$ is said to converge strongly to $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$ if

$$\lim_{n\to+\infty} \|\mathbf{A}_n\phi - \mathbf{A}\phi\|_{\mathbb{H}} = 0, \quad \forall \phi \in \mathbb{H}. \tag{1.16}$$

We will use the following monotone convergence theorem of a sequence of non-decreasing positive bounded linear operators.

Theorem 1.3.1 (Monotone convergence theorem for operators). *Let $(\mathbf{A}_n)_{n=1}^{+\infty}$ be a bounded monotone sequence of bounded linear operators on $\mathbb{H}$. Then the sequence $(\mathbf{A}_n)_{n=1}^{+\infty}$ is strongly convergent.*

Proof. Assume, for example, that $(\mathbf{A}_n)_{n=1}^{+\infty}$ is such that

$$\mathbf{A}_1 \leq \mathbf{A}_2 \leq \cdots \leq \mathbf{A}_n \leq \cdots \leq \mathbf{M}$$

for some $\mathbf{M} \in \mathfrak{B}(\mathbb{H})$. Since $\sup_n \|\mathbf{A}_n\|_\infty \leq \|\mathbf{M}\|_\infty < +\infty$, we obtain that for any $\phi \in \mathbb{H}$ the sequence $\langle \mathbf{A}_n\phi, \phi\rangle_{\mathbb{H}}$ is convergent. Therefore, due to "polarization" (1.13),

$$\begin{aligned}\langle \mathbf{A}_n\phi, \psi\rangle = \frac{1}{4}\{&\langle \mathbf{A}_n(\phi+\psi), \phi+\psi\rangle_{\mathbb{H}} - \langle \mathbf{A}_n(\phi-\psi), \phi-\psi\rangle_{\mathbb{H}} \\ &+ \imath[\langle \mathbf{A}_n(\phi+\imath\psi), \phi+\imath\psi\rangle_{\mathbb{H}} - \langle \mathbf{A}_n(\phi-\imath\psi), \phi-\imath\psi\rangle_{\mathbb{H}}]\}.\end{aligned}$$

We have that $\lim_{n\to+\infty}\langle \mathbf{A}_n\phi, \psi\rangle_{\mathbb{H}}$ exists for any $\phi, \psi \in \mathbb{H}$. Such a limit defines a self-adjoint bounded operator $\mathbf{A}$ on $\mathbb{H}$. Denote by $c = \sup_n \|\mathbf{A}_n - \mathbf{A}\|_\infty$. Then for all $\phi \in \mathbb{H}$,

$$\|\mathbf{A}_n\phi - \mathbf{A}\phi\|_{\mathbb{H}} \leq c[\langle \mathbf{A}_n\phi, \phi\rangle_{\mathbb{H}} - \langle \mathbf{A}\phi, \phi\rangle_{\mathbb{H}}] \to 0, \forall \phi \in \mathbb{H}.$$

This shows that the sequence $(\mathbf{A}_n)_{n=1}^{+\infty}$ converges strongly. This proves the monotone convergence theorem. □

Positive operators play a role similar to that of the positive real numbers in polar decomposition (see Theorem 1.8.11). Note that for any $\mathbf{T} \in \mathfrak{L}(\mathbb{H})$, the operator $\mathbf{T}^*\mathbf{T}$ is positive, since for all $\phi \in \mathbb{H}$,

$$\langle \mathbf{T}^*\mathbf{T}\phi, \phi\rangle_{\mathbb{H}} = \langle \mathbf{T}\phi, \mathbf{T}\phi\rangle_{\mathbb{H}} = \|\mathbf{T}\phi\|_{\mathbb{H}}^2 \geq 0. \tag{1.17}$$

Lemma 1.3.2. *If $\mathbf{T} \in \mathfrak{B}_+(\mathbb{H})$, then for each $\mathbf{S} \in \mathfrak{B}(\mathbb{H})$, it follows that $\mathbf{STS}^* \in \mathfrak{B}_+(\mathbb{H})$.*

Proof. Let $\mathbf{T} = \mathbf{A}^*\mathbf{A}$ for some $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$. Then

$$\mathbf{STS}^* = \mathbf{SA}^*\mathbf{AS}^* = (\mathbf{AS}^*)^*\mathbf{AS}^* \geq \mathbf{0}$$

This proves the lemma. □

Riesz proved (see Riesz [129]) that every bounded positive self-adjoint operator has a unique positive square root that has certain commutative properties below. The proof of the Riesz lemma below is omitted.

Theorem 1.3.3. *Let* $\mathbf{A}$ *be a positive self-adjoint operator in the real or complex Hilbert space* $\mathbb{H}$. *Then there exists a unique positive self-adjoint operator* $\mathbf{S}$ *such that* $\mathbf{S}^2 = \mathbf{A}$. *Furthermore,* $\mathbf{DA} \subset \mathbf{AD}$ *implies* $\mathbf{DS} \subset \mathbf{SD}$ *for any bounded linear operator* $\mathbf{D}$.

1.4 Resolvent set and spectrum

Definition 1.4.1. Let $\mathbb{X}$ be a complex Hilbert space or complex Banach space. Let $\mathbf{T} \in \mathfrak{B}(\mathbb{X})$ and $\lambda \in \mathbb{C}$. Then λ is said to be in the resolvent set $\rho(\mathbf{T})$ of $\mathbf{T}$ if the linear transformation $\mathbf{T} - \lambda\mathbf{I}$ is a bijection (i. e., one-to-one and onto), where $\mathbf{I}$ is the identity operator on $\mathbb{X}$. In this case, its (bounded) inverse $(\mathbf{T} - \lambda\mathbf{I})^{-1}$ is called the *resolvent* of $\mathbf{T}$ at λ. If $\lambda \in \sigma(\mathbf{T}) := \mathbb{C}\backslash\rho(\mathbf{T})$, then λ is in the *spectrum* $\sigma(\mathbf{T})$ of $\mathbf{T}$.

Note that if $\mathbf{T} - \lambda\mathbf{I}$ is surjective (one-to-one and onto), then the open mapping theorem implies that $(\mathbf{T}-\lambda\mathbf{I})^{-1}$ is bounded. Hence, $\lambda \in \rho(\mathbf{T})$ and both $\mathbf{T}-\lambda\mathbf{I}$ and $(\mathbf{T}-\lambda\mathbf{I})^{-1}$ are surjective bounded linear operators.

As in the finite-dimensional case, a complex number λ is called an eigenvalue of $\mathbf{T}$ if there is a nonzero vector $u \in \mathbb{H}$ such that $\mathbf{T}u = \lambda u$. In that case, $\ker(\mathbf{T} - \lambda\mathbf{I}) \neq \{0\}$, so $\mathbf{T} - \lambda\mathbf{I}$ is not one-to-one, and $\lambda \in \sigma(\mathbf{T})$. This is not the only one, however, that a complex number can belong to $\sigma(\mathbf{T})$. We further decompose the spectrum, $\sigma(\mathbf{T})$, of a bounded linear operator $\mathbf{T}$ as follows.

Definition 1.4.2 (Spectrum of bounded linear operator). Let $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$. Then

$$\sigma(\mathbf{T}) = \sigma_p(\mathbf{T}) \cup \sigma_c(\mathbf{T}) \cup \sigma_r(\mathbf{T}),$$

where:

1. $\sigma_p(\mathbf{T})$, the point spectrum of $\mathbf{T}$, consists of all $\lambda \in \sigma(\mathbf{T})$ such that $\mathbf{T}-\lambda\mathbf{I}$ is not a one-to-one map. In this case, λ is called an eigenvalue of $\mathbf{T}$ and any nonzero $u \in \mathbb{H}$ such that $\mathbf{T}u = \lambda u$ is called an eigenvector corresponding to the eigenvalue λ.
2. $\sigma_c(\mathbf{T})$, the continuous spectrum of $\mathbf{T}$, consists of those $\lambda \in \sigma(\mathbf{T})$ such that range$(\mathbf{T}-\lambda\mathbf{I})$ is a proper dense subset of $\mathbb{H}$.
3. $\sigma_r(\mathbf{T})$, the residual spectrum of $\mathbf{T}$, consists of the spectrum $\lambda \in \mathbb{C}$ that are neither an eigenvalue nor a continuous spectrum of the operator $\mathbf{T}$.

Proposition 1.4.3. *If* $\mathbf{A} : \mathbb{H} \to \mathbb{H}$ *is a self-adjoint bounded linear operator, then its point spectrum* $\sigma_p(\mathbf{A}) \subset \mathbb{R}$, *i. e., all eigenvalues of a self-adjoint operators are real.*

Proof. If $\lambda \in \sigma_p(\mathbf{A})$, then there is $0 \neq \phi \in \mathbb{H}$ such that $\mathbf{A}\phi = \lambda\phi$. In this case,

$$\lambda\langle\phi,\phi\rangle_{\mathbb{H}} = \langle\phi,\lambda\phi\rangle_{\mathbb{H}} = \langle\phi,\mathbf{A}\phi\rangle_{\mathbb{H}} = \langle\mathbf{A}\phi,\phi\rangle_{\mathbb{H}} = \langle\lambda\phi,\phi\rangle_{\mathbb{H}} = \bar{\lambda}\langle\phi,\phi\rangle_{\mathbb{H}}.$$

This shows that $\lambda = \bar{\lambda}$. Therefore, λ is a real number. □

Some of the bounded linear operators relevant to quantum communication are presented below.

1.5 Unitary operators

A unitary transformation in quantum mechanics is obtained by applying an operator $\mathbf{U}$ to a state that leaves the square modulus of the state (i. e., the probability density) unchanged. We can write this condition as the following.

Definition 1.5.1. A bounded linear operator $\mathbf{U}$ on a complex Hilbert space $\mathbb{H}$ is said to be a *unitary* operator if it satisfies the following relations:

$$\overline{\text{range}(\mathbf{U})} = \mathbb{H} \tag{1.18}$$

and

$$\langle\mathbf{U}\phi,\mathbf{U}\psi\rangle_{\mathbb{H}} = \langle\phi,\psi\rangle_{\mathbb{H}}, \quad \forall\phi,\psi \in \mathbb{H}. \tag{1.19}$$

We have the following equivalent definition of a unitary operator.

Definition 1.5.2. A unitary operator is a bounded linear operator $\mathbf{U} : \mathbb{H} \to \mathbb{H}$ on a complex Hilbert space $\mathbb{H}$ that satisfies $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I}$, where $\mathbf{U}^*$ is the adjoint of $\mathbf{U}$, and $\mathbf{I} : \mathbb{H} \to \mathbb{H}$ is the identity operator.

The weaker condition $\mathbf{U}^*\mathbf{U} = \mathbf{I}$ in the above defines an *isometry*. The other condition, $\mathbf{U}\mathbf{U}^* = \mathbf{I}$, defines a *coisometry*. Thus, a unitary operator is a bounded linear operator, which is both an isometry and a coisometry (see Section 127 of Halmos [58]) or, equivalently, a *surjective isometry* (see Proposition I.5.2 of Conway [28]).

We have the following proposition.

Proposition 1.5.3. *A linear operator* $\mathbf{U} : \mathbb{H} \to \mathbb{H}$ *is an isomorphism and* $\overline{\text{range}(\mathbf{U})} = \mathbb{H}$ *if and only if* $\mathbf{U}$ *is an unitary operator.*

To see that the above two definitions are equivalent, notice that $\mathbf{U}$ preserving the inner product implies $\mathbf{U}$ is an isometry (thus, a bounded linear operator). The fact that $\mathbf{U}$ has a dense range ensures it has a bounded inverse $\mathbf{U}^{-1}$. It is clear that $\mathbf{U}^{-1} = \mathbf{U}^*$.

If $\mathbf{U} : \mathbb{H} \to \mathbb{H}$ is a unitary operator with an eigenvalue $\lambda \in \mathbb{C}$, then (i) $\|\mathbf{U}\|_\infty = 1$ and (ii) $|\lambda| = 1$ (i. e., all eigenvalues of $\mathbf{U}$ lies on the unit circle of complex plane). This is because $\mathbf{U}^*\mathbf{U} = \mathbf{I}$ implies that $\|\mathbf{U}^*\mathbf{U}\|_\infty = \|\mathbf{U}\|_\infty^2 = \|\mathbf{I}\|_\infty = 1$. Hence, $|\mathbf{U}\|_\infty = 1$. If λ is an eigenvalue of $\mathbf{U}$, then it has an eigenvector $x \in \mathbb{H}$ with $\|x\|_\mathbb{H} = 1$ such that $\mathbf{U}x = \lambda x$. In this case, $\|\mathbf{U}x\|_\mathbb{H} = |\lambda|\|x\|_\mathbb{H}$. Since $\|\mathbf{U}\|_\infty = 1$, we have $1 = \sup_{x\neq 0} \frac{\|\mathbf{U}x\|_\mathbb{H}}{\|x\|_\mathbb{H}} = |\lambda|$. As a consequence, all eigenvalues λ of $\mathbf{U}$ are of the form $\lambda = e^{\imath\theta}, \theta \in \mathbb{R}$. In fact, on the vector space $\mathbb{C}$ of complex numbers, multiplication by a complex number of modulo 1, that is, a number of the form $e^{\imath\theta}$ for $\theta \in \mathbb{R}$, is a unitary operator, where θ is referred to as a phase, and this multiplication is referred to as multiplication by a phase. Notice that the value of θ modulo 2π does not affect the result of the multiplication, and so the independent unitary operators on $\mathbb{C}$ are parametrized by a circle.

The unitary operator $\mathbf{U}$ is a bijective and surjective map on $\mathbb{H}$. This is because $\overline{\text{range}(\mathbf{U})} = \mathbb{H}$. Therefore, $\mathbf{U}^{-1}$ (the inverse of $\mathbf{U}$) exists and $\mathbf{U}^{-1} = \mathbf{U}^*$.

Note that a unitary operator on an infinite-dimensional Hilbert space may not have eigenvalues at all.

The following are some examples of unitary operators.

Example 1.5. An $n \times n$ real matrix Q is orthogonal if $Q^\top = Q^{-1}$ and an $n \times n$ complex matrix U is unitary if $U^* = U^{-1}$, where U^* is the complex conjugate of $U^\top$.

Example 1.6. If $\mathbf{T}$ is a bounded self-adjoint operator, then

$$e^{\imath\mathbf{T}} = \sum_{n=0}^{\infty} \frac{(\imath\mathbf{T})^n}{n!}$$

is unitary, since $(e^{\imath\mathbf{T}})^* = e^{-\imath\mathbf{T}} = (e^{\imath\mathbf{T}})^{-1}$, where $\imath = \sqrt{-1}$.

1.6 Projection operators

In this section, we begin by describing some algebraic properties of a projection operator. If $\mathbb{M}$ and $\mathbb{N}$ are subspaces of a complex Hilbert space $\mathbb{H}$ such that every $\zeta \in \mathbb{H}$ can be written uniquely as $\zeta = \phi + \psi$ with $\phi \in \mathbb{M}$ and $\psi \in \mathbb{N}$, then we say that $\mathbb{H} = \mathbb{M} \oplus \mathbb{N}$ is the direct sum of $\mathbb{M}$ and $\mathbb{N}$, and we call $\mathbb{N}$ a *complementary subspace* of $\mathbb{M}$ in $\mathbb{H}$. The decomposition $\zeta = \phi + \psi$ with $\phi \in \mathbb{M}$ and $\psi \in \mathbb{N}$ is unique if and only if $\mathbb{M} \cap \mathbb{N} = \{0\}$, where 0 is the zero vector in $\mathbb{H}$. A given subspace $\mathbb{M}$ has many complementary subspaces. For example, if $\mathbb{H} = \mathbb{R}^3$ and $\mathbb{M}$ is a plane through the origin $0 \in \mathbb{R}^3$, then any line through the origin that does not lie in $\mathbb{M}$ is a complementary subspace of $\mathbb{M}$. However, every complementary subspace $\mathbb{N}$ of $\mathbb{M}$ has the same dimension, and the dimension of $\mathbb{N}$ is called the *codimension* of $\mathbb{M}$ in $\mathbb{H}$.

If $\mathbb{H} = \mathbb{M} \oplus \mathbb{N}$, then we define the projection $\mathbf{p} : \mathbb{H} \to \mathbb{M}$ of $\mathbb{H}$ onto $\mathbb{M}$ along $\mathbb{N}$ by $\mathbf{p}\zeta = \phi$, where $\zeta = \phi + \psi$ with $\phi \in \mathbb{M}$ and $\psi \in \mathbb{N}$. This projection is a bounded and linear operator on $\mathbb{H}$ with $\text{range}(\mathbf{p}) = \mathbb{M}$ and $\ker(\mathbf{p}) = \mathbb{N}$, and satisfies $\mathbf{p}^2 := \mathbf{p} \circ \mathbf{p} = \mathbf{p}$.

It can be shown that the property $\mathbf{p}^2 = \mathbf{p}$ characterizes projections. We therefore give its definition as follows.

Definition 1.6.1. A projection on the Hilbert space $\mathbb{H}$ is a bounded linear operator $\mathbf{p} : \mathbb{H} \to \mathbb{H}$ such that

$$\mathbf{p}^2 = \mathbf{p}. \tag{1.20}$$

The collection of projections on $\mathbb{H}$ will be denoted by $\mathfrak{B}_{\text{proj}}(\mathbb{H}) \subset \mathfrak{B}(\mathbb{H})$.

Any projection operator is associated with a direct sum decomposition.

Theorem 1.6.2. *Let $\mathbb{H}$ be a complex Hilbert space. The following two properties hold:*
1. *If $\mathbf{p} : \mathbb{H} \to \mathbb{H}$ is a projection, then $\mathbb{H}$ can be uniquely decomposed as $\mathbb{H} = \operatorname{range}(\mathbf{p}) \oplus \ker(\mathbf{p})$.*
2. *If $\mathbb{H} = \mathbb{M} \oplus \mathbb{N}$, where $\mathbb{M}$ and $\mathbb{N}$ are linear subspaces of $\mathbb{H}$, then there is a projection $\mathbf{p} : \mathbb{H} \to \mathbb{H}$ with $\operatorname{range}(\mathbf{p}) = \mathbb{M}$ and $\ker(\mathbf{p}) = \mathbb{N}$.*

Proof. (1). To prove (1), we first show that $\zeta \in \operatorname{range}(\mathbf{p})$ if and only if $\zeta = \mathbf{p}\zeta$. If $\zeta = \mathbf{p}\zeta$, then clearly $\zeta \in \operatorname{range}(\mathbf{p})$. On the other hand, if $\zeta \in \operatorname{range}(\mathbf{p})$, then $\zeta = \mathbf{p}\xi$ for some $\xi \in \mathbb{H}$, and since $\mathbf{p}^2 = \mathbf{p}$, it follows that $\mathbf{p}\zeta = \mathbf{p}^2\xi = \mathbf{p}\xi = \zeta$. This concludes that $\zeta \in \operatorname{rangle}(\mathbf{p})$ if and only if $\zeta = \mathbf{p}\zeta$.

Now assume that $\zeta \in \operatorname{range}(\mathbf{p}) \cap \ker(\mathbf{p})$. Then $\zeta = \mathbf{p}\zeta$ (from the conclusion from the previous paragraph) and $\mathbf{p}\zeta = \mathbf{0}$ (since $\zeta \in \ker(\mathbf{p})$). Consequently, $\zeta = \mathbf{0}$ and $\operatorname{range}(\mathbf{p}) \cap \ker(\mathbf{p}) = \{\mathbf{0}\}$. If $\zeta \in \mathbb{H}$, then we have

$$\zeta = \mathbf{p}\zeta + (\zeta - \mathbf{p}\zeta),$$

where $\mathbf{p}\zeta \in \operatorname{range}(\mathbf{p})$ and $(\zeta - \mathbf{p}\zeta) \in \ker(\mathbf{p})$, since

$$\mathbf{p}(\zeta - \mathbf{p}\zeta) = \mathbf{p}\zeta - \mathbf{p}^2\zeta = \mathbf{p}\zeta - \mathbf{p}\zeta = \mathbf{0}.$$

Thus, $\mathbb{H} = \operatorname{range}(\mathbf{p}) \oplus \ker(\mathbf{p})$. This proves (1).

(2). To prove (2), we observe that if $\mathbb{H} = \mathbb{M} \oplus \mathbb{N}$, then $\zeta \in \mathbb{H}$ has the unique decomposition $\zeta = \phi + \psi$ with $\phi \in \mathbb{M}$ and $\psi \in \mathbb{N}$, and $\mathbf{p}\zeta = \phi$ defines the required projection $\mathbf{p}$. This proves (2). □

If $S \subset \mathbb{H}$, recall that $S^\perp$ is defined by

$$S^\perp \equiv \{\phi \in \mathbb{H} \mid \langle \varphi, \phi \rangle_{\mathbb{H}} = 0 \; \forall \varphi \in S\}.$$

Let $\bigvee S$ denote the *linear span* of S, i. e., $\operatorname{span}(S)$ is the space of all finite linear combinations of elements of S. Then it is clear that $\overline{\operatorname{span}(S)}$, the *closure* of $\operatorname{span}(S)$ under the Hilbertian norm $\|\cdot\|_{\mathbb{H}}$, is the smallest closed subspace of $\mathbb{H}$, which contains S.

Let $(\mathbf{p}_i)_{i\in I}$ be a family of projections or a complex Hilbert space $\mathbb{H}$, where I is an index set. Denote by $\bigvee_{i\in I}\mathbf{p}_i$ the projection onto $\overline{\bigvee_i \mathbf{p}_i(\mathbb{H})}$, the closure of the linear subspace of $\mathbb{H}$ generated by the ranges of the $\mathbf{p}_i$'s.

Suppose $\mathbb{M}$ is a closed subspace of a Hilbert space $\mathbb{H}$. Then by Theorem 1.6.2, we have $\mathbb{H} = \mathbb{M} \oplus \mathbb{M}^{\perp}$. We call the projection of $\mathbb{H}$ onto $\mathbb{M}$ along $\mathbb{M}^{\perp}$ the *orthogonal projection* of $\mathbb{H}$ onto $\mathbb{M}$. If $\zeta = \phi+\psi$ and $\zeta' = \phi'+\psi'$, where $\phi,\phi' \in \mathbb{M}$ and $\psi,\psi' \in \mathbb{M}^{\perp}$, then the orthogonality of $\mathbb{M}$ and $\mathbb{M}^{\perp}$ implies that

$$\langle \mathbf{p}\zeta, \zeta'\rangle_{\mathbb{H}} = \langle \phi, \phi' + \psi'\rangle_{\mathbb{H}} = \langle \phi, \phi'\rangle_{\mathbb{H}} = \langle \phi+\psi, \phi'\rangle_{\mathbb{H}} = \langle \zeta, \mathbf{p}\zeta'\rangle_{\mathbb{H}}. \tag{1.21}$$

This equation states that an orthogonal projection is self-adjoint. As we will show, properties (1.20) and (1.21) characterize orthogonal projection. We therefore have the following definition of orthogonal projection.

Definition 1.6.3. An orthogonal projection on a complex Hilbert space $\mathbb{H}$ is a linear operator $\mathbf{p} : \mathbb{H} \to \mathbb{H}$ that satisfies

$$\mathbf{p}^2 = \mathbf{p} \quad \text{and} \quad \langle \mathbf{p}\zeta, \xi\rangle_{\mathbb{H}} = \langle \zeta, \mathbf{p}\xi\rangle_{\mathbb{H}} \quad \forall \zeta, \xi \in \mathbb{H}.$$

The collection of projection operators on $\mathbb{H}$ is denoted by $\mathfrak{B}_p(\mathbb{H})$ and the collection of orthogonal projection operators is denoted by $\mathfrak{B}_{\mathrm{op}}(\mathbb{H})$.

Note that it is trivial that an orthogonal projection is necessarily bounded. We therefore have the following result.

Proposition 1.6.4. *If* $\mathbf{p} : \mathbb{H} \to \mathbb{H}$ *is a nonzero orthogonal projection, then* $\|\mathbf{p}\|_\infty = 1$.

Proof. If $\zeta \in \mathbb{H}$ and $\mathbf{p}\zeta \neq 0$, then the use of the property of orthogonal projection and the Cauchy–Schwarz inequality implies that

$$\|\mathbf{p}\zeta\|_{\mathbb{H}} = \frac{\langle \mathbf{p}\zeta, \mathbf{p}\zeta\rangle_{\mathbb{H}}}{\|\mathbf{p}\zeta\|_{\mathbb{H}}} = \frac{\langle \zeta, \mathbf{p}^2\zeta\rangle_{\mathbb{H}}}{\|\mathbf{p}\zeta\|_{\mathbb{H}}} = \frac{\langle \zeta, \mathbf{p}\zeta\rangle_{\mathbb{H}}}{\|\mathbf{p}\zeta\|_{\mathbb{H}}} \leq \frac{\|\zeta\|_{\mathbb{H}}\|\mathbf{p}\zeta\|_{\mathbb{H}}}{\|\mathbf{p}\zeta\|_{\mathbb{H}}} = \|\zeta\|_{\mathbb{H}}.$$

Therefore, $\|\mathbf{p}\|_\infty \leq 1$. If $\mathbf{p} \neq \mathbf{0}$, (i. e., $\mathbf{p}$ is a nonzero operator), then there is a $\zeta \in \mathbb{H}$ with $\mathbf{p}\zeta \neq 0$, and $\|\mathbf{p}\|_\infty\|\mathbf{p}\zeta\|_{\mathbb{H}} \geq \|\mathbf{p}(\mathbf{p}\zeta)\|_{\mathbb{H}} = \|\mathbf{p}\zeta\|_{\mathbb{H}}$, so that $\|\mathbf{p}\|_\infty \geq 1$. This proves that $\|\mathbf{p}\|_\infty = 1$. □

There is a one-to-one correspondence between orthogonal projections $\mathbf{p}$ and closed subspaces $\mathbb{M}$ of $\mathbb{H}$ such that $\mathrm{range}(\mathbf{p}) = \mathbb{M}$. The kernel of the orthogonal projection is the orthogonal complement of $\mathbb{M}$.

Theorem 1.6.5. *Let* $\mathbb{H}$ *be a complex Hilbert space. The following two statements hold:*

1. *If* $\mathbf{p}$ *is an orthogonal projection on* $\mathbb{H}$, *then* $\mathrm{range}(\mathbf{p})$ *is closed, and*

$$\mathbb{H} = \mathrm{range}(\mathbf{p}) \oplus \ker(\mathbf{p})$$

is the orthogonal direct sum of $\mathrm{rangle}(\mathbf{p})$ *and* $\ker(\mathbf{p})$.

2. *If* $\mathbb{M}$ *is a closed subspace of* $\mathbb{H}$*, then there is an orthogonal projection* $\mathbf{p}$ *on* $\mathbb{H}$ *with* $\operatorname{range}(\mathbf{p}) = \mathbb{M}$ *and* $\ker(\mathbf{p}) = \mathbb{M}^{\perp}$.

Proof. (1). To prove (1), we assume that $\mathbf{p}$ is an orthogonal projection on $\mathbb{H}$. Then by Theorem 1.6.2, we have $\mathbb{H} = \operatorname{range}(\mathbf{p}) \oplus \ker(\mathbf{p})$. If $\zeta = \mathbf{p}\xi \in \operatorname{range}(\mathbf{p})$ and $\varphi \in \ker(\mathbf{p})$, then

$$\langle \zeta, \varphi \rangle_{\mathbb{H}} = \langle \mathbf{p}\xi, \varphi \rangle_{\mathbb{H}} = \langle \xi, \mathbf{p}\varphi \rangle_{\mathbb{H}} = 0,$$

so $\operatorname{range}(\mathbf{p}) \perp \ker(\mathbf{p})$. Hence, we see that $\mathbb{H}$ is the orthogonal direct sum of $\operatorname{range}(\mathbf{p})$ and $\ker(\mathbf{p})$. It follows that $\operatorname{range}(\mathbf{p}) = (\ker(\mathbf{p}))^{\perp}$, so $\operatorname{range}(\mathbf{p})$ is closed.

(2). To prove (2), we suppose that $\mathbb{M}$ is a closed subspace of $\mathbb{H}$. Then Theorem 1.6.2 implies that $\mathbb{H} = \mathbb{M} \oplus \mathbb{M}^{\perp}$. We define a projection $\mathbf{p} : \mathbb{H} \to \mathbb{H}$ by

$$\mathbf{p}\zeta = \phi, \quad \text{where } \zeta = \phi + \psi \text{ with } \phi \in \mathbb{M} \text{ and } \psi \in \mathbb{M}^{\perp}.$$

Then $\operatorname{range}(\mathbf{p}) = \mathbb{M}$, and $\ker(\mathbf{p}) = \mathbb{M}^{\perp}$. The orthogonality of $\mathbf{p}$ was shown in (1.21). □

If $\mathbf{p}$ is an orthogonal projection on $\mathbb{H}$, with $\operatorname{range}(\mathbf{p}) = \mathbb{M}$ and associated orthogonal direct sum $\mathbb{H} = \mathbb{M} \oplus \mathbb{N}$, then $\mathbf{I} - \mathbf{p}$ is the orthogonal projection with range $\mathbb{N}$ and associated orthogonal direct sum $\mathbb{H} = \mathbb{N} \oplus \mathbb{M}$.

The following are some examples of projections and/or orthogonal projections.

Example 1.7 (One-dimensional projection $\mathbf{p}_{\phi}$). Let $\phi \in \mathbb{H}$. We define the operator $\mathbf{p}_{\phi} : \mathbb{H} \to \mathbb{H}$ by

$$\mathbf{p}_{\phi}\psi = \langle \phi, \psi \rangle_{\mathbb{H}} \phi \quad \forall \psi \in \mathbb{H}.$$

This operator, called the one-dimensional projection along the vector ϕ, projects a vector orthogonally onto its component in the direction ϕ. In this case, the Hilbert space $\mathbb{H}$ has the following orthogonal direct sum decomposition:

$$\mathbb{H} = \operatorname{range}(\mathbf{p}_{\phi}) \oplus \ker(\mathbf{p}_{\phi}) = \mathbb{C}\phi \oplus (\mathbb{C}\phi)^{\perp},$$

where $\mathbb{C}\phi = \{c\phi \mid c \in \mathbb{C}\}$ is the one-dimensional subspace of $\mathbb{H}$ generated by ϕ.

Example 1.8. The space $\mathbb{H} = L^2(\mathbb{R})$ is the orthogonal direct sum of even functions $\mathbb{M}$ and odd functions $\mathbb{N}$. The orthogonal projections $\mathbf{p}$ and $\mathbf{q}$ of $\mathbb{H}$ onto $\mathbb{M}$ and $\mathbb{N}$, respectively, are given by

$$\mathbf{p}f(x) = \frac{f(x) + f(-x)}{2}, \quad \text{and} \quad \mathbf{q}f(x) = \frac{f(x) - f(-x)}{2},$$

respectively. Note that $\mathbf{I} - \mathbf{p} = \mathbf{q}$.

1.7 Finite rank and compact linear operators

1.7.1 Finite rank linear operators

A linear map, $\mathbf{T} \in \mathfrak{L}(\mathbb{H}, \mathbb{K})$, is said to be a map with rank n, if range($\mathbf{T}$) (the range of $\mathbf{T}$) is an n-dimensional subspace of $\mathbb{K}$.

One can visualize that finite-rank maps/operators are matrices (of finite size) transplanted to the infinite-dimensional setting. As such, these maps/operators may be described via linear algebra techniques. From linear algebra, we know that a rectangular matrix with complex entries, $M \in \mathbb{C}^{n\times m}$, has rank 1 if and only if the $n \times m$ matrix M is of the form

$$M = \alpha u v^*,$$

where $u = (u_1, u_2, \cdots, u_m) \in \mathbb{C}^m$ and $v = (v_1, v_2, \cdots, v_n) \in \mathbb{C}^n$ are unit vectors, $v^* = (\bar{v}_1, \bar{v}_2, \cdots, \bar{v}_n)^\top$ and $\alpha \geq 0$. Exactly the same argument shows that a linear operator $\mathbf{T}$ on a Hilbert space $\mathbb{H}$ is of rank 1 if and only if

$$\mathbf{T}\psi = \alpha \langle \psi, \phi \rangle_{\mathbb{H}} \varphi, \quad \forall \psi \in \mathbb{H},$$

where $\alpha \geq 0$, ϕ and φ are unit vectors in $\mathbb{H}$. The above representation of rank 1 operator $\mathbf{T}$ can be easily extended to the operator of rank n as follows.

If $\mathbf{T}$ is a linear operator of rank n on $\mathbb{H}$, then it has the following representation:

$$\mathbf{T}\psi = \sum_{i=1}^{n} \alpha_i \langle \psi, \phi_i \rangle_{\mathbb{H}} \varphi_i, \quad \forall \psi \in \mathbb{H},$$

where $\{\phi_i\}_{i=1}^n$ and $\{\varphi_i\}_{i=1}^n$ are orthonormal bases of range($\mathbf{T}$) $\subset \mathbb{H}$.

The linear map $\mathbf{T}$ is said of finite rank, if it is a map of rank n for some $n \in \mathbb{N}$. The collection of linear maps of finite rank from $\mathbb{H}$ to $\mathbb{K}$ will be denoted by $\mathfrak{L}_f(\mathbb{H}, \mathbb{K})$ and collection of finite rank linear operators on $\mathbb{H}$ will be denoted by $\mathfrak{L}_f(\mathbb{H})$.

It is clear that every linear map of finite rank is bounded. Therefore, $\mathfrak{L}_f(\mathbb{H}, \mathbb{K}) \subset \mathfrak{B}(\mathbb{H}, \mathbb{K})$ and $\mathfrak{L}_f(\mathbb{H}) \subset \mathfrak{B}(\mathbb{H})$.

The following result states that the collection of finite rank operators form a closed (under the operator norm $\|\cdot\|_\infty$) two-sided ideal in operator algebra (see, e. g., Rudin [134], Conway [28] and Bratteli and Robinson [15]).

Proposition 1.7.1. *If $\mathbb{H}$ is a complex Hilbert space, then $\mathfrak{L}_f(\mathbb{H})$ is a closed two-sided ideal in the Banach algebra $\mathfrak{B}(\mathbb{H})$ of bounded linear operators on $\mathbb{H}$. That is,* $\mathbf{AT} \in \mathfrak{L}_f(\mathbb{H})$ *and* $\mathbf{TA} \in \mathfrak{L}_f(\mathbb{H})$ *for all* $\mathbf{A} \in \mathfrak{L}_f(\mathbb{H})$ *and* $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$.

1.7.2 Compact linear operators

Consider a linear map $\mathbf{T} \in \mathfrak{L}(\mathbb{H}, \mathbb{K})$. Let $B \subset \mathbb{H}$ and let $\mathbf{T}(B) \subset \mathbb{K}$ denote the image of $B \subset \mathbb{H}$ under $\mathbf{T}$, i. e.,

$$\mathbf{T}(B) = \{\mathbf{T}\psi \in \mathbb{K} \mid \psi \in B \subseteq \mathbb{H}\}.$$

A linear map $\mathbf{T} \in \mathfrak{L}(\mathbb{H}, \mathbb{K})$ is said to be a compact map if $\mathbf{T}(B)$ is relatively compact in $\mathbb{K}$ (i. e., $\overline{\mathbf{T}(B)}$ (the closure of $\mathbf{T}(B)$ under $\|\cdot\|_{\mathbb{K}}$) is a compact subset of $\mathbb{K}$) for any bounded subset $B \subset \mathbb{H}$.

The collection of all compact linear maps from $\mathbb{H}$ to $\mathbb{K}$ will be denoted by $\mathfrak{L}_c(\mathbb{H}, \mathbb{K})$ and the collection of compact linear operators on $\mathbb{H}$ will be denoted by $\mathfrak{L}_c(\mathbb{H})$. It is clear that a compact linear map/operator is bounded. Therefore, $\mathfrak{L}_c(\mathbb{H}, \mathbb{K}) \subset \mathfrak{B}(\mathbb{H}, \mathbb{K})$ and $\mathfrak{L}_c(\mathbb{H}) \subset \mathfrak{B}(\mathbb{H})$.

Proposition 1.7.2. *Let $(\mathbf{T}_n)_{n=1}^{+\infty}$ be a sequence of compact linear maps from $\mathbb{H}$ to $\mathbb{K}$ that converges to $\mathbf{T}$ in the operator norm. Then $\mathbf{T}$ is a compact linear map from $\mathbb{H}$ to $\mathbb{K}$.*

Proof. Given $\epsilon > 0$, let $N > 0$ be a sufficiently large integer such that $\|\mathbf{T}_n - \mathbf{T}\|_\infty < \epsilon/2$ for all $n \geq N$. Let $B \subset \mathbb{H}$ be a bounded set. Since $\mathbf{T}_N(B)$ is relatively compact (i. e., its closure $\overline{\mathbf{T}_N(B)}$ is a compact subset of $\mathbb{K}$), there are finitely many points $y_1, y_2, \ldots, y_m$ in $\mathbb{K}$ such that for any $x \in B$ there is i such that $\|\mathbf{T}_N x - y_i\|_{\mathbb{K}} < \epsilon/2$. By the triangle inequality,

$$\begin{aligned}\|\mathbf{T}x - y_i\|_{\mathbb{K}} &\leq \|\mathbf{T}x - \mathbf{T}_n x\|_{\mathbb{K}} + \|\mathbf{T}_n x - y_i\|_{\mathbb{K}} \\ &\leq \|\mathbf{T}_n - \mathbf{T}\|_\infty \|x\|_{\mathbb{H}} + \|\mathbf{T}_n x - y_i\|_{\mathbb{K}} \\ &< \epsilon/2 + \epsilon/2 = \epsilon.\end{aligned}$$

This proves that $\mathbf{T}(B)$ is covered by finitely many balls of radius ϵ in $\mathbb{K}$. Therefore, $\overline{\mathbf{T}(B)}$ is compact in $\mathbb{K}$. This proves the proposition. □

The following result states that $\mathfrak{L}_c(\mathbb{H})$ is also a closed two-sided ideal of $\mathfrak{B}(\mathbb{H})$ just as $\mathfrak{L}_f(\mathbb{H})$ is. A proof can be found in Rudin [134] and is therefore omitted here.

Proposition 1.7.3. *If $\mathbb{H}$ is a complex Hilbert space, then $\mathfrak{L}_c(\mathbb{H})$ is a closed two-sided ideal in the Banach algebra $\mathfrak{B}(\mathbb{H})$ of bounded linear operators.*

Remark 1.1. We recall the following facts (without proofs) regarding $\mathfrak{L}_c(\mathbb{H})$ below:

1. If $\mathbf{A}$ is a compact linear operator on $\mathbb{H}$, $\lambda \in \mathbb{C}$ and $\lambda \neq 0$, then $\mathbf{A}(\mathbb{H}) = \text{rangle}(\mathbf{A})$ (the range of $\mathbf{A}$) is closed and

   $$\dim(\ker(\mathbf{A} - \lambda \mathbf{I}_{\mathbb{H}})) = \dim(\ker(\mathbf{A} - \lambda \mathbf{I}_{\mathbb{H}})^*) < \infty.$$

   This is often referred to as the Fredholm alternative theorem in functional analysis (see Conway [28], Rudin [134] and Bratteli and Robinson [15]).

2. The linear operator $\mathbf{A}$ is compact if and only if its adjoint $\mathbf{A}^*$ is compact.
3. Let $\mathbf{T}$ be a compact operator on Hilbert space $\mathbb{H}$ and let $\mathbf{A}$ be the unique positive square root of $\mathbf{T}^*\mathbf{T}$. Then (a) $\|\mathbf{A}h\|_{\mathbb{H}} = \|\mathbf{T}h\|_{\mathbb{H}}$ for all $h \in \mathbb{H}$. (b) There is a unique operator $\mathbf{U}$ such that $\|\mathbf{U}h\|_{\mathbb{H}} = \|h\|_{\mathbb{H}}$ when $h \perp \ker(\mathbf{T})$, and $\mathbf{U}h = 0$, when $h \in \ker(\mathbf{T})$ and $\mathbf{UA} = \mathbf{T}$. This is the so-called polar decomposition of a compact operator, which will be used frequently later on. A proof of the polar decomposition theorem will be provided in Theorem 1.8.11.

The proof of the following theorem can be found in Rudin [134], Conway [28] and Reed and Simon [128] and is omitted here.

Theorem 1.7.4 (Spectral theorem for self-adjoint compact operator). *Let $\mathbf{T}$ be a self-adjoint compact operator on a complex Hilbert space $\mathbb{H}$. Then:*
1. *One or the other of the two values $\pm\|\mathbf{T}\|_\infty$ is an eigenvalue of $\|\mathbf{T}\|_\infty$.*
2. $\overline{\oplus_{\lambda\in\sigma_p(\mathbf{T})}\mathbb{H}_\lambda} = \mathbb{H}$, *where $\sigma_p(\mathbf{T})$ is the point spectrum (i. e., the set of all eigenvalues) of $\mathbf{T}$ and $\mathbb{H}_\lambda$ is the subspace of $\mathbb{H}$ generated by the eigenvectors of $\mathbf{T}$ corresponding to the eigenvalue λ. That is, there is a basis of $\mathbf{T}$ that consists of eigenvectors of $\mathbf{T}$.*
3. *The eigenspaces $\mathbb{H}_\lambda$ are finite-dimensional subspaces of $\mathbb{H}$.*
4. *The only possible accumulation point of the set of eigenvalues of $\mathbf{T}$ is 0. If $\dim(\mathbb{H}) = +\infty$, then 0 is the accumulation point of the set of eigenvalues of $\mathbf{T}$.*

Corollary 1.7.5. *For every compact self-adjoint operator $\mathbf{T}$ on a complex separable infinite-dimensional Hilbert space $\mathbb{H}$, there exists a countably infinite orthonormal basis $\{e_n\}_{n=1}^{+\infty}$ of $\mathbb{H}$ consisting of eigenvectors of $\mathbf{T}$, with corresponding eigenvalues $\{\lambda_n\}_{n=1}^{+\infty} \subset \mathbb{R}$, such that $\lambda_n \to 0$.*

Example 1.9. Let $\mathbb{H} = L^2([0,1];\mathbb{C})$, the space of squared integrable complex-valued functions on $[0,1]$. The multiplication operator $\mathbf{M} : \mathbb{H} \to \mathbb{H}$ defined by

$$(\mathbf{M}f)(x) = xf(x), \quad \forall f \in \mathbb{H}, \quad x \in [0,1]$$

is a bounded self-adjoint operator on $\mathbb{H}$ that has no eigenvector, and hence, by the spectral theorem 1.7.4, cannot be compact.

Example 1.10. Let $K(x,y)$ be square integrable on the unit square $[0,1]^{\times 2} := [0,1]\times[0,1]$ and define $\mathbf{T}_K$ on $\mathbb{H} = L^2([0,1];\mathbb{C})$ by

$$(\mathbf{T}_K f)(x) = \int_0^1 K(x,y)f(y)dy.$$

Then $\mathbf{T}_K$ is a compact operator on $\mathbb{H}$. Suppose that the kernel $K(x,y)$ satisfies the Hermiticity condition

$$K(y,x) = \overline{K(x,y)}, \quad \forall x,y \in [0,1].$$

Then $\mathbf{T}_K$ is compact and self-adjoint on $\mathbb{H}$; if $\{\phi_n\}_{n=1}^{+\infty}$ is an orthonormal basis of eigenvectors, with eigenvalues $\{\lambda_n\}$, it can be proved that

$$\sum_{n=1}^{+\infty} \lambda_n^2 < \infty, \quad K(x,y) = \sum_{n=1}^{+\infty} \lambda_n \varphi_n(x)\overline{\varphi_n(y)},$$

where the sum of the series of functions is understood as L^2-convergence for the Lebesgue measure on $[0,1]^{\times 2}$.

While every bounded and closed set in a finite-dimensional space is compact (see the Heine–Borel theorem 1.7.6), the conclusion, however, is not true in infinite-dimensional spaces.

We state the Heine–Borel theorem below without proof for a later use (see, e. g., Rudin [133] and Wheeden and Zygmund [177]).

Theorem 1.7.6 (Heine–Borel theorem). *For a subset S of a finite-dimensional space $\mathbb{X}$, the following two statements are equivalent:*
1. *S is a bounded and closed subset of $\mathbb{X}$;*
2. *S is a compact subset of $\mathbb{X}$.*

Remark 1.2. A closed ball of radius $r > 0$, $B(\mathbf{0}; r) := \{\phi \in \mathbb{H} \mid \|\phi\|_{\mathbb{H}} \le r\}$, is not relatively compact if $\dim(\mathbb{H}) = \infty$. This is because if $\{e_n\}_{n=1}^{+\infty}$ is an orthonormal basis of the infinite-dimensional Hilbert space $\mathbb{H}$, then $\{re_n\}_{n=1}^{+\infty}$ is a sequence with no convergent subsequence, since all these points are distance $r\sqrt{2}$ apart, i. e., for all $i \neq j$,

$$\begin{aligned}
&\|re_i - re_j\|_{\mathbb{H}} \\
&\quad = r\|e_i - e_j\|_{\mathbb{H}} = r\sqrt{\langle e_i - e_j, e_i - e_j\rangle_{\mathbb{H}}} \\
&\quad = r\sqrt{\langle e_i, e_i\rangle_{\mathbb{H}} - \langle e_i, e_j\rangle_{\mathbb{H}} - \langle e_j, e_i\rangle_{\mathbb{H}} + \langle e_j, e_j\rangle_{\mathbb{H}}} \\
&\quad = r\sqrt{\|e_i\|_{\mathbb{H}}^2 + \|e_j\|_{\mathbb{H}}^2} \quad (\text{since } \langle e_i, e_j\rangle_{\mathbb{H}} = \langle e_j, e_i\rangle_{\mathbb{H}} = 0 \text{ for } i \neq j) \\
&\quad = r\sqrt{2}.
\end{aligned}$$

This confirms the remark.

Remark 1.3. If the Hilbert space is infinite-dimensional, then a compact operator $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$ cannot be invertible. Consequently, $\sigma(\mathbf{T})$ (the spectrum of $\mathbf{T}$) always contains 0. The spectral theorem 1.7.4 shows that $\sigma(\mathbf{T})$ consists of the eigenvalues $\{\lambda_n\}_{n=1}^{+\infty}$ of $\mathbf{T}$, and of 0 (if 0 is not already an eigenvalue). We first verify that $\mathbf{T}$ is not invertible if $\dim(\mathbb{H}) = +\infty$. If the compact operator $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$ were invertible when $\dim(\mathbb{H}) = \infty$, then $\mathbf{T}^{-1}\mathbf{T} = \mathbf{T}\mathbf{T}^{-1} = \mathbf{I}_{\mathbb{H}}$ ($\mathbf{I}_{\mathbb{H}}$ is the identity operator on $\mathbb{H}$) would be a compact operator. But $\mathbf{I}_{\mathbb{H}}(B(\mathbf{0}; 1)) = B(\mathbf{0}; 1)$, where $B(\mathbf{0}; 1) ;= \{\phi \in \mathbb{H} \mid \|\phi\|_{\mathbb{H}} \le 1\}$ is the unit closed ball in $\mathbb{H}$. When $\dim(\mathbb{H}) = \infty$, the closed unit ball $B(\mathbf{0}; 1)$ is not relatively compact (see

Remark 1.2). Therefore, $B(\mathbf{0};1)$ is not relatively compact and $\mathbf{I}_{\mathbb{H}}$ is not compact, contradicting the supposition that $\mathbf{T}$ is invertible. That is, $\mathbf{T} - \lambda\mathbf{I}_{\mathbb{H}}$ is not invertible and so $\lambda = 0$ is in $\sigma(\mathbf{T})$.

The following result gives a characterization of compact operators $\mathbf{T} \in \mathfrak{L}_c(\mathbb{H})$ (see Conway [28]).

Proposition 1.7.7. *A linear operator* $\mathbf{T}$ *on an infinite-dimensional Hilbert space* $\mathbb{H}$ *is compact if and only if it can be written in the form*

$$\mathbf{T} = \sum_{n=1}^{\infty} \lambda_n \langle \phi_n, \cdot \rangle_{\mathbb{H}} \psi_n, \tag{1.22}$$

where $\{\phi_n\}_{n=1}^{+\infty}$ *and* $\{\psi_n\}_{n=1}^{+\infty}$ *are orthonormal sets (not necessarily complete) of* $\mathbb{H}$ *and* $(\lambda_n)_{n=1}^{+\infty}$ *is a sequence of positive numbers with limit zero, called the singular values of the operator.*

It can be shown that the singular values can accumulate only at zero. If the sequence becomes stationary at zero, that is, $\lambda_{N+k} = 0$ for some $N \in \mathbb{N}$, and every $k = 1, 2, \ldots$, then the operator has finite rank, i. e., a finite-dimensional range and can be written as $T = \sum_{n=1}^{N} \lambda_n \langle \phi_n, \cdot \rangle \psi_n$. The sum on the right- hand side of (1.22) converges in the operator norm $\| \cdot \|_{\infty}$.

Now that we know $\mathfrak{L}_f(\mathbb{H}) \subset \mathfrak{B}(\mathbb{H})$ and $\mathfrak{L}_c(\mathbb{H}) \subset \mathfrak{B}(\mathbb{H})$. The following result gives the relation between $\mathfrak{L}_f(\mathbb{H})$ and $\mathfrak{L}_c(\mathbb{H})$.

Proposition 1.7.8. *Let* $\mathbb{H}$ *be a complex Hilbert space. Then* $\overline{\mathfrak{L}_f(\mathbb{H})} = \mathfrak{L}_c(\mathbb{H})$ *under the operator norm* $\| \cdot \|_{\infty}$.

1.8 Trace-class and Hilbert–Schmidt operators

1.8.1 Trace-class operators

Let $\{\phi_n\}_{n=1}^{+\infty}$ be an orthonormal basis of the complex Hilbert space $\mathbb{H}$. For each positive linear operator $\mathbf{T} \in \mathfrak{L}_+(\mathbb{H})$, define $\mathrm{tr}[\mathbf{T}]$, the trace of the operator $\mathbf{T}$, as

$$\mathrm{tr}[\mathbf{T}] := \sum_{n=1}^{+\infty} \langle \phi_n, \mathbf{T}\phi_n \rangle_{\mathbb{H}} \le +\infty. \tag{1.23}$$

Proposition 1.8.1. *The following statements hold:*

1. *The quantity* $\mathrm{tr}[\mathbf{T}]$ *is independent of the choice* $\{\phi_n\}_{n=1}^{+\infty}$ *of orthonormal basis of the Hilbert space* $\mathbb{H}$.
2. $\mathrm{tr}[\mathbf{S} + \mathbf{T}] = \mathrm{tr}[\mathbf{S}] + \mathrm{tr}[\mathbf{T}]$, *for all* $\mathbf{S}, \mathbf{T}$ *in* $\mathfrak{L}_+(\mathbb{H})$.
3. $\mathrm{tr}[c\mathbf{T}] = c\,\mathrm{tr}[\mathbf{T}]$, *for all* $c \ge 0$ *and for all* $\mathbf{T} \in \mathfrak{L}_+(\mathbb{H})$.

Proof. 1. For a given orthonormal basis $\{\phi_n\}_{n=1}^{+\infty}$, to distinguish the trace of $\mathbf{T}$ under this orthonormal basis we define $\operatorname{tr}_\phi[\mathbf{T}] = \sum_{n=1}^{+\infty}\langle \phi_n, \mathbf{T}\phi_n\rangle_{\mathbb{H}}$. Using the result (see Theorem 1.3.3) that every positive operator has a unique square root that is necessarily self-adjoint, we have

$$\operatorname{tr}_\phi[\mathbf{T}] = \sum_{n=1}^{+\infty}\langle \phi_n, \mathbf{T}\phi_n\rangle_{\mathbb{H}} = \sum_{n=1}^{+\infty}\langle \sqrt{\mathbf{T}}\phi_n, \sqrt{\mathbf{T}}\phi_n\rangle_{\mathbb{H}} = \sum_{n=1}^{+\infty}\|\sqrt{\mathbf{T}}\phi_n\|_{\mathbb{H}}^2. \tag{1.24}$$

If $\{\psi_n\}_{n=1}^{+\infty}$ is another orthonormal basis of $\mathbb{H}$, we can expand $\|\sqrt{\mathbf{T}}\phi_n\|_{\mathbb{H}}$ for each n in terms of this orthonormal basis and have

$$\begin{aligned}\operatorname{tr}_\phi[\mathbf{T}] &= \sum_{n=1}^{+\infty}\|\sqrt{\mathbf{T}}\phi_n\|_{\mathbb{H}}^2 = \sum_{n=1}^{+\infty}\left(\sum_{m=1}^{+\infty}|\langle \psi_m, \sqrt{\mathbf{T}}\phi_n\rangle_{\mathbb{H}}|^2\right)\\ &= \sum_{n=1}^{+\infty}\left(\sum_{m=1}^{+\infty}|\langle \sqrt{\mathbf{T}}\psi_m, \phi_n\rangle_{\mathbb{H}}|^2\right) \quad \text{(since } \sqrt{\mathbf{T}} \text{ is self-adjoint)}\\ &= \sum_{m=1}^{+\infty}\left(\sum_{n=1}^{+\infty}|\langle \sqrt{\mathbf{T}}\psi_m, \phi_n\rangle_{\mathbb{H}}|^2\right) = \sum_{m=1}^{+\infty}\|\sqrt{\mathbf{T}}\psi_m\|_{\mathbb{H}}^2 = \operatorname{tr}_\psi[\mathbf{T}].\end{aligned} \tag{1.25}$$

Therefore, $\operatorname{tr}[\mathbf{T}]$ defined in equation (1.23) is independent of the choice of an orthonormal basis.

Parts 2 and 3 are trivial by the definition of $\operatorname{tr}[\cdot]$. This proves the proposition. □

Given a linear operator $\mathbf{T} \in \mathfrak{L}(\mathbb{H})$, define $|\mathbf{T}| := \sqrt{\mathbf{T}^*\mathbf{T}}$, the (positive) square root of $\mathbf{T}^*\mathbf{T}$, where $\mathbf{T}^*$ is the adjoint of $\mathbf{T}$. It is clear that if $\mathbf{T} \in \mathfrak{L}(\mathbb{H})$, then for any $c \in \mathbb{C}$, $|c\mathbf{T}| = |c||\mathbf{T}|$. However, we remark that: (i) $\sqrt{\mathbf{T}^*\mathbf{T}}$ is usually not the same as $\sqrt{\mathbf{T}\mathbf{T}^*}$; (ii) it is not true in general that $|\mathbf{S}\mathbf{T}| = |\mathbf{S}||\mathbf{T}|$ or that $|\mathbf{T}| = |\mathbf{T}^*|$ and (iii) it is not true in general that $|\mathbf{S} + \mathbf{T}| \le |\mathbf{S}| + |\mathbf{T}|$.

Definition 1.8.2. $\mathbf{T} \in \mathfrak{L}(\mathbb{H})$ is a trace-class operator if $\|\mathbf{T}\|_1 := \operatorname{tr}[|\mathbf{T}|] < \infty$. The collection of all trace-class operators will be denoted by $\mathfrak{T}(\mathbb{H})$.

The map $\|\cdot\|_1 : \mathfrak{T}(\mathbb{H}) \to \mathbb{R}_+$ defined in Definition 1.8.2 is indeed a Banach norm. That is, it satisfies: (i) $\|c\mathbf{T}\|_1 = |c|\|\mathbf{T}\|_1$ for all $c \in \mathbb{C}$ and $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$; (ii) $\|\mathbf{S}+\mathbf{T}\|_1 \le \|\mathbf{S}\|_1+\|\mathbf{T}\|_1$ for all $\mathbf{S}, \mathbf{T} \in \mathfrak{T}(\mathbb{H})$; and (iii) $\|\mathbf{T}\|_1 = 0$ if and only if $\mathbf{T} = \mathbf{0} \in \mathfrak{T}(\mathbb{H})$.

It will be shown in Corollary 2.3.7 that the space of trace-class operators $\mathfrak{T}(\mathbb{H})$ is a complex Banach space under $\|\cdot\|_1$-norm.

Lemma 1.8.3. *If $\mathbf{T}$ is a self-adjoint trace-class operator, then its positive and negative parts, $\mathbf{T}_+$ and $\mathbf{T}_-$, are also trace-class operators.*

Proof. If $\mathbf{T}$ is trace class, then $\operatorname{tr}[|\mathbf{T}|] < \infty$. But $\operatorname{tr}[|\mathbf{T}|] = \operatorname{tr}[\mathbf{T}_+ + \mathbf{T}_-] = \operatorname{tr}[\mathbf{T}_+] + \operatorname{tr}[\mathbf{T}_-]$ for $\mathbf{T}_+$ and $\mathbf{T}_-$ are positive operators. Hence, $\operatorname{tr}[\mathbf{T}_+]$ and $\operatorname{tr}[\mathbf{T}_-]$ are both finite quantities. The operators $\mathbf{T}_+$ and $\mathbf{T}_-$ are thus trace class. This proves the lemma. □

The following result follow easily from the definition of trace-class operators $\mathfrak{T}(\mathbb{H})$.

Proposition 1.8.4. *The following results follow easily from the definition of trace-class operators* $\mathfrak{T}(\mathbb{H})$*:*

1. *The trace is linear functional over* $\mathfrak{T}(\mathbb{H})$*, i. e.,*

$$\mathrm{tr}[a\mathbf{S} + b\mathbf{T}] = a\,\mathrm{tr}[\mathbf{S}] + b\,\mathrm{tr}[\mathbf{T}], \quad \forall a, b \in \mathbb{C} \text{ and } \forall \mathbf{S}, \mathbf{T} \in \mathfrak{T}(\mathbb{H}).$$

2. $\mathfrak{T}(\mathbb{H})$ *forms a two-sided ideal of* $\mathfrak{B}(\mathbb{H})$*. That is, if* $\mathbf{S} \in \mathfrak{B}(\mathbb{H})$ *and* $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$*, then* $\mathbf{ST}, \mathbf{TS} \in \mathfrak{T}(\mathbb{H})$*. Furthermore,*

$$\|\mathbf{ST}\|_1 = \mathrm{tr}\big[|\mathbf{ST}|\big] \le \|\mathbf{S}\|_\infty \|\mathbf{T}\|_1 \quad \text{and} \quad \|\mathbf{TS}\|_1 = \mathrm{tr}\big[|\mathbf{TS}|\big] \le \|\mathbf{S}\|_\infty \|\mathbf{T}\|_1.$$

Moreover, $\mathrm{tr}[\mathbf{ST}] = \mathrm{tr}[\mathbf{TS}]$ *for all* $\mathbf{S} \in \mathfrak{B}(\mathbb{H})$ *and* $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$.

The proof of the following result is a consequence of Young inequality for a compact operator (see Erlijman–Farenick–Zeng [46]). The proof is rather lengthy and is omitted.

Proposition 1.8.5. *Young's inequality for trace-class operators Let* $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$*, we have*

$$\mathrm{tr}\big[|\mathbf{AB}|\big] \le \frac{\mathrm{tr}[\mathbf{A}^p]}{p} + \frac{\mathrm{tr}[\mathbf{B}^q]}{q}. \tag{1.26}$$

The equality holds if and only if for all positive real numbers p, q *satisfying* $\frac{1}{p} + \frac{1}{q} = 1$*. Moreover, the equality holds only if* $\mathbf{B}^q = \mathbf{A}^p$*.*

Proposition 1.8.6. *Let* $\mathbb{H}$ *be an infinite-dimensional complex Hilbert space. Assume that* $\mathbf{A} \in \mathfrak{T}(\mathbb{H})$ *and* $(\mathbf{T}_n)_{n=0}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$*. If* $\lim_{n\to+\infty} \|\mathbf{T}_n - \mathbf{T}_0\|_\infty = 0$ *for some* $\mathbf{T}_0 \in \mathfrak{B}(\mathbb{H})$*, then* $\lim_{n\to+\infty} \mathrm{tr}[\mathbf{T}_n\mathbf{A}] = \mathrm{tr}[\mathbf{T}_0\mathbf{A}]$*.*

Proof. We first note from Part 2 of Proposition 1.8.4 that $\mathrm{tr}[\mathbf{T}_n\mathbf{A}] = \mathrm{tr}[\mathbf{A}\mathbf{T}_n] < +\infty$ and $\mathrm{tr}[\mathbf{T}_0\mathbf{A}] = \mathrm{tr}[\mathbf{A}\mathbf{T}_0] < +\infty$. As $\mathbf{T}_n$ converges to $\mathbf{T}_0$ under the operator norm $\|\cdot\|_\infty$, there is a constant $d > 0$ such that $\sup_n \|\mathbf{T}_n\|_\infty \le d$. For any $\epsilon > 0$, there exists a finite rank projection $\mathbf{P}_\epsilon$ such that $\|\mathbf{A} - \mathbf{P}_\epsilon\mathbf{A}\mathbf{P}_\epsilon\|_1 < \epsilon/(2d+1)$ because $\mathbf{A} \in \mathfrak{T}(\mathbb{H})$. Since $\mathbf{P}_\epsilon$ is of finite rank, this together with the fact that $\lim_{n\to+\infty} \|\mathbf{T}_n - \mathbf{T}_0\|_\infty = 0$, imply that

$$\lim_{n\to+\infty} \big\|\mathbf{P}_\epsilon(\mathbf{T}_n - \mathbf{T}_0)\mathbf{P}_\epsilon\big\|_\infty = 0.$$

So, for above $\epsilon > 0$, there exists some N such that

$$\big\|\mathbf{P}_\epsilon(\mathbf{T}_n - \mathbf{T}_0)\mathbf{P}_\epsilon\big\|_\infty < \frac{\epsilon}{(2d+1)\|\mathbf{A}\|_1}$$

whenever $n > N$. Thus, we have

$$\begin{aligned}\left|\mathrm{tr}\left[(\mathbf{T}_n - \mathbf{T}_0)\mathbf{A}\right]\right| &\le \left|\mathrm{tr}\left[(\mathbf{T}_n - \mathbf{T}_0)(\mathbf{A} - \mathbf{P}_\epsilon \mathbf{A} \mathbf{P}_\epsilon)\right]\right| + \left|\mathrm{tr}\left[(\mathbf{T}_n - \mathbf{T}_0)\mathbf{P}_\epsilon \mathbf{A} \mathbf{P}_\epsilon\right]\right| \\ &\le \|\mathbf{T}_n - \mathbf{T}_0\|_\infty \|\mathbf{A} - \mathbf{P}_\epsilon \mathbf{A} \mathbf{P}_\epsilon\|_1 + \left\|\mathbf{P}_\epsilon(\mathbf{T}_n - \mathbf{T}_0)\mathbf{P}_\epsilon\right\|_\infty \|\mathbf{A}\|_1 \\ &< 2d\|\mathbf{A} - \mathbf{P}_\epsilon \mathbf{A} \mathbf{P}_\epsilon\|_1 + \frac{\epsilon}{2d+1} < \frac{2d\epsilon}{2d+1} + \frac{\epsilon}{2d+1} = \epsilon.\end{aligned}$$

Therefore, $\lim_{n\to+\infty} \mathrm{tr}[\mathbf{T}_n\mathbf{A}] = \mathrm{tr}[\mathbf{T}_0\mathbf{A}]$. This proves the proposition. □

Lemma 1.8.7. *Let* $\mathbf{A}_k, \mathbf{A} \in \mathfrak{B}(\mathbb{H})$ *such that*

$$\sup_k \|\mathbf{A}_k\|_\infty < +\infty \quad \text{and} \quad \lim_{k\to+\infty} \|\mathbf{A}_k\phi - \mathbf{A}\phi\|_{\mathbb{H}} = 0, \quad \forall \phi \in \mathbb{H}.$$

Let $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$. *Then*

$$\lim_{k\to+\infty} \|\mathbf{A}_k\mathbf{T} - \mathbf{A}\mathbf{T}\|_1 = 0 \quad \text{and} \quad \lim_{k\to+\infty} \|\mathbf{T}\mathbf{A}_k - \mathbf{T}\mathbf{A}\|_1 = 0.$$

Proof. We only need to prove $\lim_{k\to+\infty} \|\mathbf{A}_k\mathbf{T} - \mathbf{A}\mathbf{T}\|_1 = 0$. The proof of the second conclusion is similar. We first note $\sup_k \|\mathbf{A}_k\|_\infty < +\infty$ and $\lim_{k\to+\infty} \|\mathbf{A}_k\phi - \mathbf{A}\phi\|_{\mathbb{H}} = 0$ for all $\phi \in \mathbb{H}$ imply that $\lim_{k\to+\infty} \|\mathbf{A}_k - \mathbf{A}\|_\infty = 0$. In this case,

$$\lim_{k\to+\infty} \|\mathbf{A}_k\mathbf{T} - \mathbf{A}\mathbf{T}\|_1 \le \lim_{k\to+\infty} \|\mathbf{A}_k - \mathbf{A}\|_\infty \|\mathbf{T}\|_1 = 0.$$

This proves the lemma. □

Corollary 1.8.8. *If* $(\mathbf{P}_k)_{k=1}^{+\infty}$ *is a sequence of projectors on* $\mathbb{H}$ *that converges in the strong operator topology to the identity, and if* $\rho \in \mathfrak{T}_+(\mathbb{H})$, *then*

$$\lim_{k\to+\infty} \|\mathbf{P}_k\rho\mathbf{P}_k - \rho\|_1 = 0.$$

It turns out that there is a simple connection between trace-class operators and compact operators. The following result states that every trace-class operator is compact and every compact operator is a trace-class operator if all of its singular values (see (1.22) for the definition of singular values) are finite.

The proof of the following theorem can be found in Reed and Simon [128] and is omitted here.

Theorem 1.8.9. *Every* $\mathbf{A} \in \mathfrak{T}(\mathbb{H})$ *is compact. Conversely, a compact operator* $\mathbf{A}$ *is in* $\mathfrak{T}(\mathbb{H})$ *if and only if* $\sum_i \lambda_i < +\infty$, *where* $\{\lambda_i\}$ *are the singular values of* $\mathbf{A}$.

Denote $\mathfrak{L}_{c,\lambda<+\infty}(\mathbb{H}) \subset \mathfrak{L}_c(\mathbb{H})$ as

$$\mathfrak{L}_{c,\lambda<+\infty}(\mathbb{H}) = \{\mathbf{A} \in \mathfrak{L}_c(\mathbb{H}) \mid \text{all singular values are finite}\}.$$

The following corollary serves to provide one with an intuitive understanding of the size of $\mathfrak{T}(\mathbb{H})$, and thus allows for the application of approximation arguments in what follows.

Corollary 1.8.10. *The finite rank operators are dense in $\mathfrak{T}(\mathbb{H})$ under the $\|\cdot\|_1$-norm. That is, $\overline{\mathfrak{L}_f(\mathbb{H})}^{\|\cdot\|_1} = \mathfrak{T}(\mathbb{H})$, where $\overline{\{\cdots\}}^{\|\cdot\|_1}$ denotes the closure of $\{\cdots\}$ under the $\|\cdot\|_1$.*

Summarizing from all our presentations on all different classes of bounded linear operators such as projection operators, finite rank operators, compact operators and trace-class operators, we have the following relationship among them:

$$\mathfrak{L}_f(\mathbb{H}) \subset \overline{\mathfrak{L}_f(\mathbb{H})}^{\|\cdot\|_1} = \mathfrak{T}(\mathbb{H}) = \mathfrak{L}_{c,\lambda<+\infty}(\mathbb{H}) \subset \mathfrak{L}_c(\mathbb{H}) = \overline{\mathfrak{L}_f(\mathbb{H})}^{\|\cdot\|_\infty} \subset \mathfrak{B}(\mathbb{H}).$$

We have following duality relation between $\mathfrak{T}(\mathbb{H})$ and $\mathfrak{B}(\mathbb{H})$ through $\langle\langle\cdot,\cdot\rangle\rangle : \mathfrak{T}(\mathbb{H})\times \mathfrak{B}(\mathbb{H}) \to \mathbb{C}$ as

$$\langle\langle \rho, \mathbf{A}\rangle\rangle := \mathrm{tr}[\rho\mathbf{A}], \quad \forall \rho \in \mathcal{S}(\mathbb{H}) \text{ and } \forall \mathbf{A} \in \mathfrak{B}(\mathbb{H}). \tag{1.27}$$

An element $\rho \in \mathcal{S}(\mathbb{H})$ is called a quantum state (see Definition 2.4.1 and Chapter 3 for a definition of quantum state and its properties), which is the counterpart in probability distribution in classical/noncommutative probability theory. A positive bounded linear operator $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$ represents an observable in the quantum system. The quantity $\langle\langle \rho, \mathbf{A}\rangle\rangle := \mathrm{tr}[\rho\mathbf{A}]$ represents the expectation of the observable $\mathbf{A}$ when the quantum system is under the quantum state ρ.

1.8.2 Hilbert–Schmidt operator

Given an orthonormal basis $(\phi_n)_{n=1}^{+\infty}$ of the complex Hilbert space $\mathbb{H}$ and a bounded linear operator $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$, we put the Hilbert–Schmidt norm $\|\mathbf{T}\|_{\mathrm{HS}}$ as follows:

$$\|\mathbf{T}\|_{\mathrm{HS}} := \left(\sum_{n=1}^{+\infty} \|\mathbf{T}\phi_n\|_{\mathbb{H}}^2\right)^{1/2} \leq +\infty. \tag{1.28}$$

It can be shown that $\|\mathbf{T}\|_{\mathrm{HS}} = \|\mathbf{T}^*\|_{\mathrm{HS}}$. An operator $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$ is said to be a Hilbert–Schmidt operator if it has a finite Hilbert–Schmidt norm (i. e., $\|\mathbf{T}\|_{\mathrm{HS}} < +\infty$). The space of all Hilbert–Schmidt operators will be denoted by $\mathfrak{HS}(\mathbb{H})$. Note that $\mathfrak{HS}(\mathbb{H})$ is itself a Hilbert space under the Hilbert–Schmidt inner product $\langle\cdot,\cdot\rangle_{\mathrm{HS}} : \mathfrak{HS}(\mathbb{H})\times\mathfrak{HS}(\mathbb{H}) \to \mathbb{C}$ defined by

$$\langle \mathbf{S}, \mathbf{T}\rangle_{\mathrm{HS}} = \sum_n \langle \mathbf{S}\phi_n, \mathbf{T}\phi_n\rangle_{\mathbb{H}}, \quad \mathbf{S}, \mathbf{T} \in \mathfrak{HS}(\mathbb{H}).$$

It appears that both the Hilbert–Schmidt norm $\|\cdot\|_{\mathrm{HS}}$ and the trace defined above are expressed in terms of an orthonormal basis $(\phi_n)_{n=1}^{+\infty}$. However, a further analysis (which we shall omit here) shows that they are both independent of the orthonormal basis chosen.

For a finite-dimensional Hilbert space $\mathbb{H} = \mathbb{C}^n$, the polar decomposition of an $n \times n$ real or complex matrix $\mathbf{A}$ is a factorization of the form $\mathbf{A} = \mathbf{UP}$, where $\mathbf{U}$ is a unitary matrix (a rotation matrix) and $\mathbf{P}$ is a positive-semidefinite Hermitian matrix (a scaling of the space along a set of n orthogonal axes), both of size $n \times n$. The polar decomposition of a square matrix $\mathbf{A}$ always exists. In fact, if $\mathbf{A}$ is invertible, the decomposition is unique, and the factor $\mathbf{P}$ will be positive-definite. In that case, $\mathbf{A}$ can be written uniquely in the form $\mathbf{A} = \mathbf{U}e^{\mathbf{X}}$, where $\mathbf{U}$ is unitary and $\mathbf{X}$ is the unique self-adjoint logarithm of the matrix $\mathbf{P}$. The polar decomposition can also be defined as $\mathbf{A} = \mathbf{PU}$ where $\mathbf{P}$ and $\mathbf{U}$ have the same properties as above (but are different matrices, in general, for the same $\mathbf{A}$).

The polar decomposition of a matrix can be seen as the matrix analog of the polar form of a complex number as $z = ur$, where r is its absolute value (a nonnegative real number), and u is a complex number with unit norm (an element of the circle group).

For the infinite-dimensional Hilbert space $\mathbb{H}$, the polar decomposition of a bounded linear operator $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$ can roughly be stated as follows. If $\mathbf{A}$ is a bounded linear operator, then there is a unique factorization of $\mathbf{A}$ as a product $\mathbf{A} = \mathbf{UP}$ where $\mathbf{U}$ is a partial isometry, $\mathbf{P}$ is a nonnegative self-adjoint operator and the initial space of $\mathbf{U}$ is the closure of the range of $\mathbf{P}$.

A bounded linear operator $\mathbf{U} \in \mathfrak{B}(\mathbb{H})$ is said to be an isometry if $\|\mathbf{U}\phi\|_{\mathbb{H}} = \|\phi\|_{\mathbb{H}}$ for all $\phi \in \mathbb{H}$, a *partial isometry* if its restriction to $(\ker(\mathbf{U}))^{\perp}$ is an isometry.

The *polar decomposition* is formally stated and proved as follows (see, e. g., Hall [56]).

Theorem 1.8.11 (Polar decomposition). *Let* $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$. *Then there exists a partial isometry* $\mathbf{U}$ *such that* $\mathbf{A} = \mathbf{U}|\mathbf{A}|$, *where* $|\mathbf{A}| = \sqrt{\mathbf{A}^*\mathbf{A}}$. *The operator* $\mathbf{U}$ *is uniquely determined by the condition that* $\ker(\mathbf{U}) = \ker(\mathbf{A})$.

Proof. Let $\mathbf{U} : \text{range}(|\mathbf{A}|) \to \text{range}(\mathbf{A})$ such that $\mathbf{U}(|\mathbf{A}|\phi) = \mathbf{A}\phi$ for all $\phi \in \mathbb{H}$. This implies that

$$\|\mathbf{U}(|\mathbf{A}|\phi)\|_{\mathbb{H}} = \|\mathbf{A}\phi\|_{\mathbb{H}}, \quad \forall \phi \in \mathbb{H}.$$

Then

$$\begin{aligned}\||\mathbf{A}|\phi\|_{\mathbb{H}}^2 &= \langle |\mathbf{A}|\phi, |\mathbf{A}|\phi\rangle_{\mathbb{H}} \\ &= \langle \phi, |\mathbf{A}|^2\phi\rangle_{\mathbb{H}} \quad (\text{since } |\mathbf{A}|^* = |\mathbf{A}|) \\ &= \langle \phi, \mathbf{A}^*\mathbf{A}\phi\rangle_{\mathbb{H}} = \langle \mathbf{A}\phi, \mathbf{A}\phi\rangle_{\mathbb{H}} = \|\mathbf{A}\phi\|_{\mathbb{H}}^2.\end{aligned}$$

The definition of $\mathbf{U}$ and the above equality together imply that

$$\|\mathbf{U}(|\mathbf{A}|\phi)\|_{\mathbb{H}} = \|\mathbf{A}\phi\|_{\mathbb{H}} = \||\mathbf{A}|\phi\|_{\mathbb{H}}, \quad \forall \phi \in \text{range}(\mathbf{A}).$$

Consequently, $\mathbf{U}$ is an isometry on range($|\mathbf{A}|$) = range($\mathbf{A}$). Clearly, $|\mathbf{A}|\phi = 0$ is equivalent to $\mathbf{A}\phi = 0$ and, therefore, $\ker(|\mathbf{A}|) = \ker(\mathbf{A})$, and thus, $\ker(\mathbf{U}) = \ker(\mathbf{A})$. Therefore, the restriction of $\mathbf{U}$ to $(\ker(\mathbf{U}))^{\perp}$ (= $(\ker(\mathbf{A}))^{\perp}$ = range($\mathbf{A}$) by Theorem 1.2.6) is an isometry, and hence, $\mathbf{U}$ is a partial isometry. This proves the polar decomposition theorem. □

The properties of polar decomposition is used in the proof of the following theorem.

Theorem 1.8.12. *A bounded linear operator is a product of two Hilbert–Schmidt operators if and only if it is a trace-class operator.*

Proof. (⇐) Let us assume $\mathrm{tr}[|\mathbf{a}|] < +\infty$ (i. e., $\mathbf{a} \in \mathfrak{T}(\mathbb{H})$) and let $\mathbf{b} = \sqrt{|\mathbf{a}|}$. We note that $\mathbf{b} = \sqrt{|\mathbf{a}|}$ is self-adjoint by Theorem 1.3.3 and claim that $\mathbf{b}$ is a Hilbert–Schmidt operator. To prove the claim, we choose any orthonormal basis $(\phi_n)_{n=1}^{+\infty}$ of $\mathbb{H}$ and have

$$\begin{aligned}\|\mathbf{b}\|_{\mathrm{HS}}^2 &= \sum_{n=1}^{+\infty} \langle \mathbf{b}\phi_n, \mathbf{b}\phi_n \rangle_{\mathbb{H}} = \sum_{n=1}^{+\infty} \langle \phi_n, \mathbf{b}^*\mathbf{b}\phi_n \rangle_{\mathbb{H}} \\ &= \sum_{n=1}^{+\infty} \langle \phi_n, \mathbf{b}^2\phi_n \rangle_{\mathbb{H}} = \sum_{n=1}^{+\infty} \langle \phi_n, |\mathbf{a}|\phi_n \rangle_{\mathbb{H}} = \mathrm{tr}[|\mathbf{a}|] < +\infty.\end{aligned}$$

In this case, by Theorem 1.8.11, $\mathbf{a} = \mathbf{u}|\mathbf{a}| = (\mathbf{u}\mathbf{b})\mathbf{b}$ is a product of two Hilbert–Schmidt operators.

(⇒) Conversely, let $\mathbf{a} = \mathbf{h}\mathbf{k}$ be a product of two Hilbert–Schmidt operators $\mathbf{h}$ and $\mathbf{k}$ on $\mathbb{H}$. Then $|\mathbf{a}| = \mathbf{u}^*\mathbf{a} = (\mathbf{u}^*\mathbf{h})\mathbf{k}$ and gives

$$\mathrm{tr}[|\mathbf{a}|] \le \|\mathbf{u}^*\mathbf{h}\|_{\mathrm{HS}}\|\mathbf{k}\|_{\mathrm{HS}} \le \|\mathbf{h}\|_{\mathrm{HS}}\|\mathbf{k}\|_{\mathrm{HS}} < +\infty,$$

since $\|\mathbf{u}\|_{\infty} \le 1$. Therefore, $\mathbf{a}$ is a trace-class operator. This proves the theorem. □

2 Formulation of quantum systems

In this chapter, we give a concise and yet rigorous mathematical formulation of a generic infinite-dimensional quantum system based on the following set of postulates originated from von Neumann [172] (see also Chang [22–24]). These postulates are commonly accepted by quantum probabilists and quantum physicists alike as the starting point for a systematic study of quantum systems.

Postulate 1. With every *quantum system*, there is associated a separable complex Hilbert space $\mathbb{H}$ on which a C^* or von Neumann algebras of linear operators $\mathcal{A}$ is defined. This complex Hilbert space $\mathbb{H}$ is called in physics terminology the *space of states*. The Hilbert space of a *composite quantum system* can be represented as a tensor product of Hilbert spaces of the component systems involved.

Postulate 2. Given a C^* or von Neumann algebra of operators $\mathcal{A}$ on $\mathbb{H}$ for the quantum system, the space of quantum states $\mathcal{S}(\mathcal{A})$ of the quantum system then consists of all positive trace-class operators $\rho \in \mathcal{A}$ with unit trace, $\mathrm{tr}[\rho] = 1$. The *pure states* are projection operators onto one-dimensional subspaces of $\mathbb{H}$. A state ρ will be called the *density operator* or *density matrix* if $\mathrm{tr}[\rho\mathbf{a}] = \mathrm{tr}[\mathbf{a}]$ for all $\mathbf{a} \in \mathcal{A}$.

Postulate 3. An observable of the quantum system is a positive operator-valued measure $\mathbf{a}$ defined on a certain Borel measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Specifically, for each Borel set $B \in \mathcal{B}(\mathbb{R})$, $\mathbf{a}(B)$ is a self-adjoint linear (but not necessarily bounded) operator on the Hilbert space $\mathbb{H}$.

Postulate 4. A process of measurement in a quantum system is the correspondence between the observable-state pair $(\mathbf{a}, \rho)$ and the probability measure $\mu_{\mathbf{a}}$ on the real Borel measurable space. For every Borel subset $E \in \mathcal{B}(\mathbb{R})$, the quantity $0 \le \mu_{\mathbf{a}}(E) \le 1$ is the probability that when a quantum system is in the state ρ, the result of the measurement of the observable $\mathbf{a}$ belongs to E. The expectation value (the mean value) of the observable $\mathbf{a}$ in the state ρ is

$$\langle \mathbf{a}|\rho\rangle = \int_{-\infty}^{\infty} \lambda d\mu_{\mathbf{a}}(\lambda),$$

where $\mu_{\mathbf{a}}(\lambda) = \mu_{\mathbf{a}}((-\infty, \lambda))$ is the distribution function for the probability measure $\mu_{\mathbf{a}}$.

While these postulates are widely-known in the quantum physics community, a detailed and rigorous mathematical formulation for each of the above postulates for infinite-dimensional quantum systems is needed. This will be beneficial to all readers, especially to those mathematical inclined readers who are not familiar with quantum physics and to those who desire to place quantum physics in a rigorous mathematical framework. In addition, the topics in this chapter will serve as the foundation upon which the subsequent chapters are built. Interested readers are also referred to Chang

https://doi.org/10.1515/9783110788105-002

[24], Brattellie and Robinson [15] and Takesaki [168] for a further reading in this area. Unless otherwise stated, we shall assume that $\mathbb{H}$ is an infinite- dimensional (separable) complex Hilbert space (and hence the quantum system it represents is an infinite-dimensional system), unless otherwise stated. The notation, definitions and the preliminary functional analytic results introduced or outlined in Chapter 1 are applicable in this chapter in order to describe the quantum system $\mathbb{H}$ mathematically.

Let $\mathbb{H}$ be a complex Hilbert space that presents a quantum system. The notation, definitions and the preliminary functional analytic results introduced or outlined in Chapter 1 are applicable in this chapter in order to describe the quantum system $\mathbb{H}$ mathematically.

2.1 Operator topologies

This section is devoted to the studies of some operator topologies, including strong/norm, weak, σ-strong, σ-weak topologies that play an important role in describing a quantum system. Details of these treatments can be found in Chang [22–24] and Bratteli and Robinson [15].

We first recall that the operator norm $\|\mathbf{a}\|_\infty$ for a bounded linear operator $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$ is defined by

$$\|\mathbf{a}\|_\infty := \sup_{\phi \neq \mathbf{0}} \frac{\|\mathbf{a}(\phi)\|_{\mathbb{H}}}{\|\phi\|_{\mathbb{H}}} = \sup_{\|\phi\|_{\mathbb{H}}=1} \|\mathbf{a}(\phi)\|_{\mathbb{H}}, \quad \forall \mathbf{a} \in \mathfrak{B}(\mathbb{H}).$$

In the norm topology, all open sets are generated by the open set of the following form:

$$\{\mathbf{a} \in \mathfrak{B}(\mathbb{H}) \mid \|\mathbf{a}\|_\infty < \epsilon\}.$$

Let $\mathbf{a}, \mathbf{b} \in \mathfrak{B}(\mathbb{H})$; we say that $\mathbf{a} \leq \mathbf{b}$ if $\mathbf{b} - \mathbf{a} \geq \mathbf{0}$. Recall that a bounded linear operator $\mathbf{a}$ is a positive operator (denoted by $\mathbf{a} \geq \mathbf{0}$) if there exist a $\mathbf{b} \in \mathfrak{B}(\mathbb{H})$ such that $\mathbf{a} = \mathbf{b}^*\mathbf{b} := \mathbf{b}^* \circ \mathbf{b}$.

The following definitions and remark can be found in Chang [23].

Definition 2.1.1. A subset of operators $(\mathbf{a}_\alpha)_{\alpha \in L} \subset \mathfrak{B}(\mathbb{H})$ is said to be a net if L is a totally ordered set with the order relation "≤." The net $(\mathbf{a}_\alpha)_{\alpha \in L}$ (or simply $(\mathbf{a}_\alpha)_\alpha$ if there is no danger of ambiguity) is said to be increasing if $\alpha \leq \tilde{\alpha}$ implies that $\mathbf{a}_\alpha \leq \mathbf{a}_{\tilde{\alpha}}$. The increasing net $(\mathbf{a}_\alpha)_\alpha$ is said to be bounded above if there exists an $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$ such that

$$\mathbf{a}_\alpha \leq \mathbf{a}, \quad \forall \alpha \in L.$$

Definition 2.1.2. Let $(\mathbf{a}_\alpha)_{\alpha \in L}$ be a net of operators in $\mathfrak{B}(\mathbb{H})$ and let $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$. We say that:

1. $(\mathbf{a}_\alpha)_{\alpha\in L}$ converges to $\mathbf{a}$ in norm-topology if $(\|\mathbf{a}_\alpha - \mathbf{a}\|_\infty)_\alpha$ converges to 0.
2. $(\mathbf{a}_\alpha)_{\alpha\in L}$ converges strongly to $\mathbf{a}$ if $\mathbf{a}_\alpha(\phi)$ converges to $\mathbf{a}(\phi)$ in $\|\cdot\|_{\mathbb{H}}$-norm for every $\phi \in \mathbb{H}$.
3. $(\mathbf{a}_\alpha)_{\alpha\in L}$ converges to $\mathbf{a}$ in strong* topology if the net $(\|\mathbf{a}_\alpha(\zeta)\|_{\mathbb{H}} + \|\mathbf{a}_\alpha^*(\zeta)\|_{\mathbb{H}})_\alpha$ converges to $\|\mathbf{a}(\zeta)\|_{\mathbb{H}} + \|\mathbf{a}^*(\zeta)\|_{\mathbb{H}}$ for all $\zeta \in \mathbb{H}$.
4. $(\mathbf{a}_\alpha)_{\alpha\in L}$ converges weakly to $\mathbf{a}$ if $\langle\varphi, \mathbf{a}_\alpha(\phi)\rangle_{\mathbb{H}}$ converges to $\langle\varphi, \mathbf{a}(\phi)\rangle_{\mathbb{H}}$ for every $\phi, \varphi \in \mathbb{H}$.
5. $(\mathbf{a}_\alpha)_{\alpha\in L}$ converges σ-strongly to $\mathbf{a}$ if the sum $\sum_{n=1}^\infty \|(\mathbf{a}_\alpha - \mathbf{a})\phi_n\|_{\mathbb{H}}^2$ converges to 0 for every sequences $(\phi_n)_n$ in $\mathbb{H}$ such that $\sum_{n=1}^\infty \|\phi_n\|_{\mathbb{H}}^2 < \infty$.
6. $(\mathbf{a}_\alpha)_{\alpha\in L}$ converges σ-weakly to $\mathbf{a}$ if the sum $\sum_{n=1}^\infty \langle\varphi_n, \mathbf{a}_\alpha(\phi_n)\rangle_{\mathbb{H}}$ converges to $\sum_{n=1}^\infty \langle\varphi_n, \mathbf{a}(\phi_n)\rangle_{\mathbb{H}}$ for every pair of sequences $(\phi_n)_n$, $(\varphi_n)_n$ in $\mathbb{H}$ such that $\sum_{n=1}^\infty \|\phi_n\|_{\mathbb{H}}^2 < \infty$ and $\sum_{n=1}^\infty \|\varphi_n\|_{\mathbb{H}}^2 < \infty$.
7. The increasing net $(\mathbf{a}_\alpha)_{\alpha\in L}$ converges to $\mathbf{a}$ in weak* topology if for every trace-class operator $\mathbf{b} \in \mathfrak{T}(\mathbb{H})$ the sequence $(\mathrm{tr}[\mathbf{b}\mathbf{a}_\alpha])_\alpha$ converges to $\mathrm{tr}[\mathbf{b}\mathbf{a}]$.

The following remark can be found in Chang [23].

Remark 2.1. Let $(\mathbf{a}_\alpha)_\alpha$ be an increasing net of operators in $\mathfrak{B}(\mathbb{H})$, and let $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$. We have the following observations:

1. The increasing net sequence $(\mathbf{a}_\alpha)_\alpha$ converges strongly to $\mathbf{a}$ if and only if it converges in operator norm, i. e., $(\|\mathbf{a}_\alpha - \mathbf{a}\|_\infty)_\alpha$ converges to 0. This is because $(\mathbf{a}_\alpha)_\alpha$ converges to $\mathbf{a}$ in operator norm if and only if $(\|(\mathbf{a}_n - \mathbf{a})\phi\|_{\mathbb{H}})_\alpha$ converges to 0 for all $\phi \in \mathbb{H}$.
2. $(\mathbf{a}_\alpha)_\alpha$ converges σ-weakly to $\mathbf{a}$ if and only if it converges to $\mathbf{a}$ in weak* topology. The weak* *topology* (or *σ-weak topology*) is the weak topology arising from the operator pairing $(\mathbf{a}, \mathbf{b}) \mapsto \mathrm{tr}[\mathbf{ab}]$ of $\mathfrak{B}(\mathbb{H})$ with the trace-class operators. That is, it is the weakest topology that makes the map $\mathbf{a} \mapsto \mathrm{tr}[\mathbf{ab}]$ continuous for the all trace-class operator $\mathbf{b} \in \mathfrak{T}(\mathbb{H})$.
3. The weak* topology is finer than the weak operator topology but the weak and weak* topology coincide on bounded subsets of $\mathbb{H}$.
4. The positive cone $\mathfrak{B}_+(\mathbb{H})$ is sequentially closed in the weak operator topology, i. e., for $(\mathbf{T}_k)_{k=1}^{+\infty} \subset \mathfrak{B}_+(\mathbb{H})$ such that $(\mathbf{T}_k)_{k=1}^{+\infty}$ converge to $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$ in the weak operator topology. It follows that $\mathbf{T} \geq \mathbf{0}$.

We recall Zorn's lemma (see Zorn [184], Rudin [133, 134] and Reed and Simon [128]) as follows. A totally ordered subset A of a partially ordered set S with an ordering inequality $\preceq$ will be called a chain. The chain A is said to have an upper bound if there exists a $u \in S$ such that $x \preceq u$ for all $x \in A$. The element $m \in S$ is said to be a maximal element for S if $m \preceq x$ implies $x = m$.

Zorn's lemma is stated below without proof.

Lemma 2.1.3 (Zorn's lemma). *Suppose a partially ordered set S has the property that every chain has an upper bound. Then the set S contains at least one maximal element.*

The following result, which can be considered as a special case of Zorn's lemma (Lemma 2.1.3), shows that if the increasing net $(\mathbf{a}_\alpha)_{\alpha\in L} \subset \mathfrak{B}_+(\mathbb{H})$ is bounded above, then it has the least upper bound that belongs to $\mathfrak{B}_+(\mathbb{H})$.

The following result can be found in Chang [23].

Proposition 2.1.4. *Let $(\mathbf{a}_\alpha)_{\alpha\in L}$ be an increasing net in $\mathfrak{B}_+(\mathbb{H})$ with an upper bound in $\mathfrak{B}_+(\mathbb{H})$. Then the net $(\mathbf{a}_\alpha)_{\alpha\in L}$ has a least upper bound (l. u. b) to be denoted by $\vee_{\alpha\in L}\mathbf{a}_\alpha$. In this case, the net converges σ-strongly to $\vee_{\alpha\in L}\mathbf{a}_\alpha$.*

Proof. Let A_α be the weak closure of the set $\{\mathbf{a}_\beta \mid \beta > \alpha\}$. Since the closed unit ball,

$$B(\mathbf{0};1) := \{\mathbf{a} \in \mathfrak{B}(\mathbb{H}) \mid \|\mathbf{a}\|_\infty \le 1\} \subset \mathfrak{B}(\mathbb{H}),$$

is weakly compact by the Banach–Alouglu theorem (see Theorem 1.1.4), there exists an element $\mathbf{a}$ in $\cap_\alpha A_\alpha$. For all $\mathbf{a}_\alpha$, the set

$$\{\mathbf{c} \in \mathfrak{B}_+(\mathbb{H}) \mid \mathbf{c} \ge \mathbf{a}_\alpha\}$$

is σ-weakly closed and contains A_α; hence, $\mathbf{a} \ge \mathbf{a}_\alpha$. Consequently, $\mathbf{a} \ge \mathbf{a}_\alpha$ for all α and lies in the weak closure of $\{\mathbf{a}_\alpha \mid \alpha \in L\}$. If $\mathbf{b}$ is another operator majorizing $\{\mathbf{a}_\alpha \mid \alpha \in L\}$, then it majorizes its weak closure; thus, $\mathbf{a} \ge \mathbf{a}$ and $\mathbf{a} := \vee_\alpha \mathbf{a}_\alpha$ is the least upper bound of $\{\mathbf{a}_\alpha \mid \alpha \in L\}$. Finally, if $\zeta \in \mathbb{H}$ then

$$\begin{aligned} \|(\mathbf{a}-\mathbf{a}_\alpha)\zeta\|_{\mathbb{H}}^2 &\le \|\mathbf{a}-\mathbf{a}_\alpha\|_\infty \|(\mathbf{a}-\mathbf{a}_\alpha)^{1/2}\zeta\|_{\mathbb{H}}^2 \\ &\le \|\mathbf{a}\|_\infty \langle \zeta, (\mathbf{a}-\mathbf{a}_\alpha)\zeta\rangle_{\mathbb{H}} \\ &\to 0 \quad \text{as } \alpha \text{ goes to } +\infty. \end{aligned}$$

This proves the proposition. □

2.2 Operator algebras

This section is devoted to studies of some algebras of operators on $\mathbb{H}$, including C^*-algebra and von Neumann algebra, that play an important role in describing a quantum system.

As usual, the addition $\mathbf{a}+\mathbf{b}$, the multiplication $\mathbf{ab}$ and the involution $\mathbf{a}^*$ on $\mathfrak{L}(\mathbb{H})$ are defined as follows. For all $\mathbf{a}, \mathbf{b} \in \mathfrak{L}(\mathbb{H})$, $\phi \in \mathbb{H}$ and $a, b \in \mathbb{C}$:

1. $(\mathbf{a}+\mathbf{b})(\phi) = \mathbf{a}(\phi) + \mathbf{b}(\phi)$;
2. $(\mathbf{ab})(\phi) = \mathbf{a}(\mathbf{b}(\phi)) = (\mathbf{a} \circ \mathbf{b})(\phi)$;
3. $(a\mathbf{a} + b\mathbf{b})^* = \bar{a}\mathbf{a}^* + \bar{b}\mathbf{b}^*$, $(\mathbf{a}^*)^* = \mathbf{a}$, and $(\mathbf{ab})^* = \mathbf{b}^*\mathbf{a}^*$.

2.2.1 C^*-algebras

Definition 2.2.1. A complex algebra $\mathcal{A}$ is a vector space of operators $\mathcal{A}$ over $\mathbb{C}$, which is closed under operator addition "+" (i. e., $\mathbf{a}+\mathbf{b} \in \mathcal{A}$ for all $\mathbf{a}, \mathbf{b} \in \mathcal{A}$) and under vector multiplication/convolution "$\circ$" (i. e., $\mathbf{ab} := \mathbf{a} \circ \mathbf{b} \in \mathcal{A}$ for all $\mathbf{a}, \mathbf{b} \in \mathcal{A}$) and satisfies the following distributive and associative laws:
1. $(a\mathbf{a} + b\mathbf{b})\mathbf{c} = a\mathbf{ac} + b\mathbf{bc}$ and $\mathbf{c}(a\mathbf{a} + b\mathbf{b}) = a\mathbf{ca} + b\mathbf{cb}$;
2. $\mathbf{a}(\mathbf{bc}) = (\mathbf{ab})\mathbf{c}$; for all $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ in $\mathcal{A}$ and a, b in $\mathbb{C}$.

Definition 2.2.2. A Banach algebra is a complex Banach space of operators $\mathcal{A}$ under a norm $\|\cdot\|$, where $\mathcal{A}$ is a complex algebra of operators (see Definition 2.2.1) and the norm $\|\cdot\|$ is a Banach norm that satisfies the inequality $\|\mathbf{ab}\| \le \|\mathbf{a}\|\|\mathbf{b}\|$ for all $\mathbf{a}$ and $\mathbf{b}$ in $\mathcal{A}$.

Recall that $\|\cdot\| : \mathcal{A} \to \mathbb{R}_+$ is said to be a Banach norm if (i) $\|\mathbf{a}\| \ge 0$ for all $\mathbf{a} \in \mathcal{A}$; (ii) $\|c\mathbf{a}\| = |c|\|\mathbf{a}\|$ for all $c \in \mathbb{C}$ and $\mathbf{a} \in \mathcal{A}$; (iii) $\|\mathbf{a} + \mathbf{b}\| \le \|\mathbf{a}\| + \|\mathbf{b}\|$ and (iv) $\|\mathbf{a}\| = 0$ if and only if $\mathbf{a} = \mathbf{0}$.

Definition 2.2.3. Let $\mathcal{A}$ be a complex algebra of operators. A map $\mathbf{a} \mapsto \mathbf{a}^*$ is called an involution on $\mathcal{A}$ if it satisfies:
1. $(\mathbf{a}^*)^* = \mathbf{a}$;
2. $\mathbf{ab}^* = \mathbf{b}^*\mathbf{a}^*$;
3. $(a\mathbf{a} + b\mathbf{b})^* = \bar{a}\mathbf{a}^* + \bar{b}\mathbf{b}^*$,

for all $\mathbf{a}$ and $\mathbf{b}$ in $\mathcal{A}$, and all a, b in $\mathbb{C}$. A complex algebra equipped with an involution * is called a *-algebra.

It is clear that the adjointness of an operator $\mathbf{a} \mapsto \mathbf{a}^*$ defined is an involution as defined in Definition 2.2.3. For convenience, we shall treat $\mathbf{a}^*$, the involution of $\mathbf{a}$, as the adjoint of the operator $\mathbf{a} \in \mathfrak{L}(\mathbb{H})$.

Definition 2.2.4. A C^*-algebra of bounded linear operators on $\mathbb{H}$ is a Banach algebra $\mathcal{A}$ under the operator norm $\|\cdot\|_\infty$ (see (1.11)) equipped with an involution "*" (which is taken to be the adjointness) if it satisfies the condition

$$\|\mathbf{a}^*\mathbf{a}\|_\infty = \|\mathbf{a}\|_\infty^2 \quad \forall \mathbf{a} \in \mathcal{A}. \tag{2.1}$$

Let $\mathcal{A}$ be a C^*-algebra of bounded linear operators on $\mathbb{H}$. Then $\mathcal{A}$ is said to be:
1. unital if it contains the unit operator $\mathbf{1}_\mathbb{H}$, where $\mathbf{1}_\mathbb{H}\mathbf{a} = \mathbf{a}\mathbf{1}_\mathbb{H} = \mathbf{a}$ for all $\mathbf{a} \in \mathcal{A}$. Note that $\mathbf{1}_\mathbb{H} = \mathbf{1}_\mathbb{H}^*$ with $\|\mathbf{1}_\mathbb{H}\|_\infty = \|\mathbf{1}_\mathbb{H}^*\|_\infty = 1$.
2. Abelian if $\mathbf{ab} = \mathbf{ba}$ for all $\mathbf{a}, \mathbf{b} \in \mathcal{A}$.

We assume throughout the end of this book that all C^*-algebras are unital because it can expanded to a larger C^*-algebra that includes the identity operator (see Arveson

[2] for such detail). Therefore, due to the remark above we can and will assume all C^*-algebras involved are unital throughout the end of this book.

The following are some examples of C^*-algebras (see Chang [22, 23]):

1. First, it is clear that both $\mathfrak{L}(\mathbb{H})$ and $\mathfrak{B}(\mathbb{H})$ are complex vector spaces of operators under pointwise addition and scalar multiplication (by complex numbers) and they are complex algebras under additional operation of composition (i. e., $\mathbf{ab} \equiv \mathbf{a} \circ \mathbf{b} \equiv \mathbf{a}(\mathbf{b})$ for all $\mathbf{a}, \mathbf{b} \in \mathfrak{L}(\mathbb{H})$). Adding the involution $\mathbf{a} \mapsto \mathbf{a}^*$ of taking adjoints, both $\mathfrak{L}(\mathbb{H})$ and $\mathfrak{B}(\mathbb{H})$ are *-algebras. Furthermore, $\mathfrak{B}(\mathbb{H})$ is a C^*-algebra, since it is complete with respect to the operator norm $\|\cdot\|_\infty$ and satisfies the basic axiom
$$\|\mathbf{a}^*\mathbf{a}\|_\infty = \|\mathbf{a}\|_\infty^2, \quad \forall \mathbf{a} \in \mathfrak{B}(\mathbb{H}).$$
To prove the above assertion, we note that $\mathfrak{B}(\mathbb{H})$ contains the identity operator $\mathbf{I}_\mathbb{H}$ and adjoint * as an involution, and it is certainly complete with respect to the operator norm $\|\cdot\|_\infty$, since $\mathfrak{B}(\mathbb{H})$ is a Banach space under this operator norm. It remains to show that
$$\|\mathbf{a}^*\mathbf{a}\|_\infty = \|\mathbf{a}\|_\infty^2, \quad \forall \mathbf{a} \in \mathfrak{B}(\mathbb{H}).$$
We first note that $\|\mathbf{a}^*\mathbf{a}\|_\infty \leq \|\mathbf{a}\|_\infty^2$, since $\|\mathbf{a}^*\mathbf{a}\|_\infty \leq \|\mathbf{a}^*\|_\infty\|\mathbf{a}\|_\infty = \|\mathbf{a}\|_\infty^2$. On the other hand,
$$\|\mathbf{a}\|_\infty^2 = \sup_{\|\phi\|_\mathbb{H} \leq 1} \|\mathbf{a}(\phi)\|_\mathbb{H}^2 = \sup_{\|\phi\|_\mathbb{H} \leq 1} \langle \mathbf{a}\phi, \mathbf{a}\phi\rangle_\mathbb{H} = \sup_{\|\phi\|_\mathbb{H} \leq 1} \langle \mathbf{a}^*\mathbf{a}\phi, \phi\rangle_\mathbb{H} \leq \|\mathbf{a}^*\mathbf{a}\|_\infty.$$
Since $\mathbf{a}$ is bounded so are $\mathbf{a}^*$ and $\mathbf{a}^*\mathbf{a}$. Therefore, $\|\mathbf{a}^*\mathbf{a}(\phi)\|_\mathbb{H} \leq \|\mathbf{a}^*\mathbf{a}\|_\infty\|\phi\|_\mathbb{H}$ for all $\phi \in \mathbb{H}$ and the last inequality of the above expression holds true.
2. C^*-algebras are noncommutative analogues of the algebra $C(K)$ of continuous complex-valued functions over a compact space K with uniform topology, the complex conjugation as involution, and the constant function 1 as unit. Thus, bounded linear functionals on C^*-algebras are the non-commutative analogues of (bounded) complex measures on a compact space K.
3. Let $\mathcal{A}$ be a C^*-algebra with unit $\mathbf{1}$ and let $\mathbf{a}, \mathbf{b} \in \mathcal{A}$.
   (1) If $\mathbf{a}$ is self-adjoint then $\mathbf{a} \leq \|\mathbf{a}\|_\infty \mathbf{1}$.
   (2) If $\mathbf{0} \leq \mathbf{a} \leq \mathbf{b}$, then $\|\mathbf{a}\|_\infty \leq \|\mathbf{b}\|_\infty$.
   (3) If $\mathbf{a}, \mathbf{b} \in \mathcal{A}_+$, then $\|\mathbf{a} - \mathbf{b}\|_\infty \leq \max\{\|\mathbf{a}\|_\infty, \|\mathbf{b}\|_\infty\}$.

Definition 2.2.5. A bounded linear map $\pi : \mathcal{A} \to \mathcal{B}$ between two C^*-algebras $\mathcal{A} \subset \mathfrak{B}(\mathbb{H})$ and $\mathcal{B} \subset \mathfrak{B}(\mathbb{K})$ is called a *-homomorphism, if for any $\mathbf{a}, \mathbf{b} \in \mathcal{A}$ and $a, b \in \mathbb{C}$,
$$\pi(a\mathbf{a} + b\mathbf{b}) = a\pi(\mathbf{a}) + b\pi(\mathbf{b}),$$
$$\pi(\mathbf{ab}) = \pi(\mathbf{a})\pi(\mathbf{b}),$$

and

$$\pi(\mathbf{a}^*) = \pi(\mathbf{a})^*.$$

The following remark and definition can be found in Chang [23].

Remark 2.2. It can be shown that any $*$-homomorphism π between two C^*-algebras is nonexpansive, i. e., $\|\pi\|_{\mathcal{A},\mathcal{B}} := \sup_{\|\mathbf{a}\|_\infty=1} \|\pi(\mathbf{a})\|_\infty \leq 1$. Therefore, a $*$-homomorphism between $*$-algebras is isometry due to the norm condition.

Definition 2.2.6. Let $\mathcal{A}$ and $\mathcal{B}$ be two C^*-algebras of bounded linear operators (but not necessarily acting on the same Hilbert space). A bijective (i. e., one-to-one and onto) *-homomorphism $\pi : \mathcal{A} \to \mathcal{B}$ is called a *-isomorphism, in which case $\mathcal{A}$ and $\mathcal{B}$ are called isomorphic. A *-isomorphism is a *-automorphism if it maps a C^*-algebra onto itself.

2.2.2 von Neumann algebras

For each subset $\mathcal{A} \subset \mathfrak{B}(\mathbb{H})$, define $\mathcal{A}'$, the commutant of $\mathcal{A}$, as

$$\mathcal{A}' = \{\mathbf{b} \in \mathfrak{B}(\mathbb{H}) \mid \mathbf{ab} = \mathbf{ba} \quad \forall \mathbf{a} \in \mathcal{A}\}.$$

Clearly, $\mathcal{A}'$ is a complex algebra of operators containing the identity operator $\mathbf{I}_{\mathbb{H}}$. If $\mathcal{A}$ is invariant under the *-operation, i. e., if $\mathbf{a} \in \mathcal{A}$ implies that $\mathbf{a}^* \in \mathcal{A}$, then the commutant $\mathcal{A}'$ is a C^*-algebra of bounded linear operators acting on $\mathbb{H}$, which is closed under any of the topologies defined in Definition 2.1.2.

As shown in Chang [22], there are at least two equivalent ways to define a von Neumann algebra. The first and most common way is to define them as weakly closed *-algebras of bounded linear operators (on a Hilbert space) containing the identity $\mathbf{I}$. In this definition, the weak (operator) topology can be replaced by many other common operator topologies. The *-algebras of bounded linear operators that are closed in the norm topology are C^*-algebras, so in particular any von Neumann algebra is a C^*-algebra.

We present the first definition of von Neumann algebra below (see Takesaki [168], Dixmier [39], Bratteli and Robinson [15] Sakai [137] and Chang [22–24]).

Definition 2.2.7 (von Neumann algebra 1). A von Neumann algebra $\mathcal{A}$ is a *-subalgebra of $\mathfrak{B}(\mathbb{H})$ that contains the identity operator $\mathbf{I}$ and is strongly closed. That is, if

$$\mathbf{a}_i \in \mathcal{A} \text{ and } \lim_{i\to\infty} \mathbf{a}_i(\phi) = \mathbf{a}(\phi) \quad \forall \phi \in \mathbb{H},$$

then $\mathbf{a} \in \mathcal{A}$.

Note that a subalgebra $\mathcal{A}$ of $\mathfrak{B}(\mathbb{H})$ is called a sub*-algebra if it is invariant under the involution *. That is,

$$\mathbf{a}^* \in \mathcal{A} \text{ for all } \mathbf{a} \in \mathcal{A}.$$

The second definition is that a von Neumann algebra is a subset of the bounded linear operators that is closed under the involution $*$ and equal to its double commutant, or equivalently

Definition 2.2.8 (von Neumann algebra 2). A von Neumann algebra of bounded linear operators on $\mathbb{H}$ is a *-subalgebra $\mathcal{A}$ of $\mathfrak{B}(\mathbb{H})$ such that $\mathcal{A} = \mathcal{A}''$.

The von Neumann double commutant theorem (von Neumann [172]) states that these two definitions are equivalent. A proof is included here for the sake of completeness and can be skipped from the first reading.

Theorem 2.2.9 (von Neumann bicommutant theorem). . *Let $\mathcal{A}$ be a* sub**-algebra of* $\mathfrak{B}(\mathbb{H})$ *containing the identity operator* $\mathbf{I}_{\mathbb{H}}$. *Then the bicommutant $\mathcal{A}''$ is exactly the strong (= weak) closure of $\mathcal{A}$. In particular, $\mathcal{A}$ is a von Neumann algebra if and only if $\mathcal{A} = \mathcal{A}''$.*

Proof. It is clear that the closure of $\mathcal{A}$ under strong convergence (see Definition 2.1.2) is contained in $\mathcal{A}''$. We must prove conversely that, for every $\mathbf{a} \in \mathcal{A}''$, every $\epsilon > 0$ and every finite family $x_1, \ldots, x_n$ of elements of $\mathbb{H}$, there exists $\mathbf{b} \in \mathcal{A}$ such that $\|\mathbf{b}x_i - \mathbf{a}x_i\|_{\mathbb{H}} \le \epsilon$ for all i. We begin with the case of one single vector x. Let $\mathbb{K} = \{\mathbf{b}x \mid \mathbf{b} \in \mathcal{A}\}$ be the closed subspace in $\mathbb{H}$ generated by $\mathbf{b}x$, $\mathbf{b} \in \mathcal{A}$. In this case, $\mathbb{K}$ is closed under the action of $\mathcal{A}$, and the same is true for $\mathbb{K}^\perp$. That is, $\mathcal{A}(\mathbb{K}) \subseteq \mathbb{K}$ and $\mathcal{A}(\mathbb{K}^\perp) \subset \mathbb{K}^\perp$. This is because if $\langle y, \mathbf{b}x\rangle_{\mathbb{H}} = 0$ for every $\mathbf{b} \in \mathcal{A}$, we also have for $\mathbf{c} \in \mathcal{A}$ $\langle \mathbf{c}y, \mathbf{b}x\rangle_{\mathbb{H}} = \langle y, \mathbf{c}^*\mathbf{b}x\rangle_{\mathbb{H}} = 0$. This means that the orthogonal projection $\mathbf{p}$ on $\mathbb{K}$ commutes with every $\mathbf{b} \in \mathcal{A}$. That is, $\mathbf{p} \in \mathcal{A}'$. Then $\mathbf{a} \in \mathcal{A}''$ commutes with $\mathbf{p}$, so that $\mathbf{a}x \in \mathbb{K}$. This implies that there exist a sequence $(\mathbf{a}_n)_{n\ge1}$ in $\mathcal{A}$ such that $\lim_{n\to\infty} \|\mathbf{a}_n x - \mathbf{a}x\|_{\mathbb{H}} = 0$. We now extend the above result to n vectors. Let $\hat{\mathbb{H}}$ be the direct sum of n copies of $\mathbb{H}$. A linear operator on $\hat{\mathbb{H}}$ can be considered as a $n \times n$ matrix of operators on $\mathbb{H}$, and we call $\hat{\mathcal{A}}$ the algebra consisting of all operators $\hat{\mathbf{a}}$ repeating $\mathbf{a} \in \mathcal{A}$ along the diagonal ($\hat{\mathbf{a}}(y_1 \oplus \ldots \oplus y_n) = \mathbf{a}y_1 \oplus \ldots \mathbf{a}y_n$). It is easy to see that $\hat{\mathcal{A}}$ is a von Neumann whose commutant consists of all matrices with coefficients in $\mathcal{A}'$, and whose bicommutant consists of all diagonals $\hat{\mathbf{a}}$ with $\mathbf{a} \in \mathcal{A}''$. Then the approximation result for n vectors $(x_1, \ldots, x_n)$ in $\mathcal{A}$ follows from the preceding result applied to $x_1 \oplus \ldots \oplus x_n$. This proves the theorem. □

Remark 2.3. As mentioned in Chang [23] that if $\mathbb{H}$ is finite-dimensional (say $\mathbb{H} = \mathbb{C}^N$ for some positive integer N), then every *-subalgebra $\mathcal{A}$ of $\mathfrak{B}(\mathbb{H})$ is also a von Nuemann algebra. This is because in finite-dimensional space $\mathbb{H}$, normal topology and Euclidean topology coincide. However, there are *-algebras that are not a von Nuemann algebra as shown in the following example, when $\mathbb{H}$ is infinite-dimensional.

The following example can be found in Chang [23].

Example 2.1. Let $\mathbb{H} = L^2([0,1])$, the space of Lebesque square integrable functions defined on the interval $[0,1]$ and let $\mathcal{A} = C([0,1])$, the commutative algebra of continuous functions on the unit interval. We can consider $A \in \mathcal{A}$ as an operator on $\mathbb{H}$ under pointwise multiplication, i. e., $(A\psi)(x) = A(x)\psi(x)$ for every $\psi \in \mathbb{H}$. Then $\mathcal{A}$ is a *-algebra but is not closed under normal topology, and hence it is not a von Neumann algebra. Indeed, we can construct a sequence of increasing functions $\{A_n\}_{n=2}^{\infty} \subset \mathcal{A}$ that converges pointwise to the indicator function $\chi_{[0,1/2]} \notin \mathcal{A}$. For example, we can let $A_n \in \mathcal{A}$ be defined by

$$A_n(x) = \begin{cases} 1 & \text{for } x \in [0, \frac{1}{2} - \frac{1}{n}), \\ -n(x - \frac{1}{2} + \frac{1}{n}) + 1 & \text{for } x \in [\frac{1}{2} - \frac{1}{n}, \frac{1}{2}), \\ 0 & \text{for } x \in [\frac{1}{2}, 1]. \end{cases}$$

It is a well-known fact that a net $(\mathbf{a}_\alpha)_{\alpha \in L}$ in $\mathcal{A}$ is σ-weakly convergent to $\mathbf{a} \in \mathcal{A}$ if and only if $tr[\rho(\mathbf{a}_n - \mathbf{a})] \to 0$ as $n \to +\infty$ for all $\rho \in \mathfrak{T}(\mathbb{H})$.

Definition 2.2.10. Given any $\mathcal{S} \subset \mathfrak{B}(\mathbb{H})$ that is invariant under the involution "$*$", we call $vN(\mathcal{S}) = \mathcal{S}''$ the von Neumann algebra generated by $\mathcal{S}$.

This construction of a von Neumann subalgebra $vN(\mathcal{S})$ has been used frequently in treatment of quantum stochastics and control (see, e. g., Chang [24]).

Definition 2.2.11. Let $\mathcal{A}$ be a von Neumann algebra of bounded linear operators on some Hilbert space $\mathbb{H}$. Let $\mathbf{T}$ be a self-adjoint operator defined on a dense domain $\mathcal{D}(\mathbf{T})$ of $\mathbb{H}$. $\mathbf{T}$ is said to be *affiliated* to $\mathcal{A}$ if $(\mathbf{T} - \imath\mathbf{I})^{-1}$ is an element of $\mathcal{A}$, where $\imath = \sqrt{-1}$.

Let $\mathcal{A}_{sa}$ be the subset of self-adjoint operators in $\mathcal{A}$.

Definition 2.2.12. Let $\mathcal{A}$ and $\mathcal{B}$ be von Neumann subalgebras of $\mathfrak{B}(\mathbb{H})$ and $\mathfrak{B}(\mathbb{K})$, respectively. A positive linear map $\Upsilon : \mathcal{A} \to \mathcal{B}$ is said to be normal if, for each bounded monotone increasing net $(\mathbf{a}_\alpha)_{\alpha \in L}$ in $\mathcal{A}_{sa}$ with $\mathbf{a} = \vee_{\alpha \in L} \mathbf{a}_\alpha$, the net $(\Upsilon(\mathbf{a}_\alpha))_{\alpha \in L}$ increases to $\vee_{\alpha \in L} \Upsilon(\mathbf{a}_\alpha) = \Upsilon(\mathbf{a}) \in \mathcal{B}_{sa}$.

Definition 2.2.13. A linear map $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ is called *ultra-weakly continuous* if

$$\lim_{n \to \infty} \operatorname{tr}[\rho \Upsilon(\mathbf{X}_n)] = \operatorname{tr}[\rho \Upsilon(\mathbf{X})]$$

whenever

$$\lim_{n \to \infty} \operatorname{tr}[\rho \mathbf{X}_n] = \operatorname{tr}[\rho \mathbf{X}] \quad \text{for all trace-class operator } \rho \text{ in } \mathfrak{B}(\mathbb{H}).$$

2.3 Bounded linear functionals

Let $\mathbb{X}$ be a Banach or Hilbert space over the complex field $\mathbb{C}$.

1. A function $\rho : \mathbb{X} \to \mathbb{C}$ is said to be a *linear functional* if

$$\rho(ax + by) = a\rho(x) + b\rho(y), \quad \forall a, b \in \mathbb{C} \text{ and } \forall x, y \in \mathbb{X}.$$

2. The linear functional $\rho : \mathbb{X} \to \mathbb{C}$ is said to be bounded (or continuous) if there exists a constant $K > 0$ such that

$$\big|\rho(x)\big| \leq K\|x\|_{\mathbb{X}}, \quad \forall x \in \mathbb{X}.$$

3. A function $\rho : \mathbb{X} \to \mathbb{R}$ is said to be a *sublinear functional* if

$$\rho(ax) = a\rho(x) \text{ and } \rho(x + y) \leq \rho(x) + \rho(y), \quad \forall a \in \mathbb{R}_+ \text{ and } \forall x, y \in \mathbb{X}.$$

Note from Section 1.1 that every seminorm (in particular, every norm) $\|\cdot\|_{\mathbb{X}} : \mathbb{X} \to \mathbb{R}$ is a sublinear functional since

$$\|cx\|_{\mathbb{X}} = |c|\|x\|_{\mathbb{X}}, \quad \forall c \in \mathbb{C} \text{ and } \forall x \in \mathbb{X},$$
$$\|x + y\|_{\mathbb{X}} \leq \|x\|_{\mathbb{X}} + \|y\|_{\mathbb{X}}, \quad \forall x, y \in \mathbb{X}.$$

If $\rho : \mathbb{X} \to \mathbb{C}$ is a bounded linear functional, we define the (operator) norm

$$\|\rho\|_\infty = \sup_{x \neq 0} \frac{|\rho(x)|}{\|x\|_{\mathbb{X}}} = \sup_{\|x\|_{\mathbb{X}}=1} \big|\rho(x)\big|. \tag{2.2}$$

The space of bounded linear functionals on $\mathbb{X}$ will be denoted by $\mathbb{X}^*$, which will be referred to as the *topological dual space* of $\mathbb{X}$.

Let $\mathbb{X} = \mathbb{H}$. The space $\mathbb{H}^*$, the space of bounded linear functionals on $\mathbb{H}$, is often referred to as the topological dual space of $\mathbb{H}$ and can be identified with $\mathbb{H}$ through the following celebrated Riesz representation theorem due originally to Frechet [51] and Riesz [129].

Theorem 2.3.1 (Riesz representation theorem). *The map* $\mathbf{F} : \mathbb{H} \to \mathbb{C}$ *is a bounded linear functional on a complex Hilbert space* $\mathbb{H}$ *if and only if there is a unique* $\psi \in \mathbb{H}$ *such that*

$$\mathbf{F}(\phi) = \langle \psi, \phi \rangle_{\mathbb{H}}, \quad \forall \phi \in \mathbb{H}.$$

In this case, $\|\mathbf{F}\|_\infty = \|\psi\|_{\mathbb{H}}$, *where* $\|\mathbf{F}\|_\infty$ *is the operator norm of* $\mathbf{F}$ *defined in (2.2).*

The Riesz representation theorem implies that $\mathbb{H}^*$, the topological dual of $\mathbb{H}$, can be identified with the Hilbert space $\mathbb{H}$ itself. We frequently use Dirac's notation $\langle \psi|_{\mathbb{H}}$:

$\mathbb{H} \to \mathbb{C}$ for a bounded linear functional, where $\psi \in \mathbb{H}$ is the Riesz representation of $\mathbf{F} \in \mathbb{H}^*$ and write

$$\mathbf{F}(\phi) = \langle \psi | \phi \rangle_{\mathbb{H}} := \langle \psi, \phi \rangle_{\mathbb{H}}, \quad \forall \phi \in \mathbb{H}.$$

For notational simplicity, we often write $\langle \psi |_{\mathbb{H}}$ as $\langle \psi |$ and $\langle \psi | \phi \rangle_{\mathbb{H}}$ as $\langle \psi | \phi \rangle$ when there is no danger of ambiguity.

If $\mathbb{D} \subset \mathbb{H}$ is dense and if $\mathbf{F} : \mathbb{D} \to \mathbb{C}$ is a bounded linear functional, then $\mathbf{F}$ can be extended as a bounded linear functional on $\mathbb{H}$. This is because if $\phi \in \mathbb{H}$, then there exists a sequence $(\phi_n)_{n=1}^{+\infty}$ in $\mathbb{D}$ such that $(s) \lim_{n\to+\infty} \phi_n = \phi$ (i. e., $\lim_{n\to+\infty} \|\phi_n - \phi\|_{\mathbb{H}} = 0$). In this case, we define its extension $\mathbf{F} : \mathbb{H} \to \mathbb{C}$ by $\mathbf{F}(\phi) = \lim_{n\to+\infty} \mathbf{F}(\phi_n)$. One can easily prove the extended $\mathbf{F} : \mathbb{H} \to \mathbb{C}$ is a bounded linear functional.

The following *uniform boundedness theorem* (or *Banach–Steinhaus theorem*) is one of the fundamental results in functional analysis. In its basic form, it asserts that for a family of bounded linear functionals on a Banach space, pointwise boundedness is equivalent to uniform boundedness in operator norm. The theorem was first published in Banach and Steinhaus [4]. The proof is omitted here.

Theorem 2.3.2 (Uniform boundedness theorem). *Let $\mathbb{X}$ be a complete normed linear space equipped with the norm $\|\cdot\|_{\mathbb{X}}$, where $\mathbb{X} = \mathbb{B}$ or $\mathbb{H}$. Suppose that $(\mathbf{F}_n)_{n=1}^{+\infty}$ is a sequence of linear functionals on $\mathbb{X}$ such that the set $\{\mathbf{F}_n(u)\}_{n=1}^{+\infty}$ is bounded in $\mathbb{C}$ for each $u \in \mathbb{X}$. Then the sequence $(\|\mathbf{F}_n\|_\infty)_{n=1}^{+\infty}$ is bounded in $\mathbb{R}$, where $\|\mathbf{F}_n\|_\infty$ is the operator norm of $\mathbf{F}_n$ defined in (2.2).*

We recall one of the important tools, the Hahn–Banach theorem, on bounded linear functional $\mathbb{X}^*$ below. The Hahn–Banach theorem is a central tool in functional analysis. It allows the extension of bounded linear functionals defined on a subspace of some vector space to the whole space, and it also shows that there are "enough" continuous linear functionals defined on every normed vector space to make the study of the dual space "interesting." Another version of the Hahn–Banach theorem is known as the hyperplane separation theorem, and has numerous uses in convex geometry.

Let us recall the Hahn–Banach theorem (in a general form) as follows (see Theorem 3.2 of Rudin [134]).

Theorem 2.3.3 (Hahn–Banach theorem). *If $\mathcal{N} : \mathbb{X} \to \mathbb{R}$ is a sublinear functional on the real vector space $\mathbb{X}$ and $\varphi : \mathbb{U} \to \mathbb{R}$ is a linear functional on a linear subspace $\mathbb{U}$ of $\mathbb{X}$, which is dominated by $\mathcal{N}$ on $\mathbb{U}$,*

$$\varphi(x) \le \mathcal{N}(x), \quad \forall x \in \mathbb{U}$$

then there exists a linear extension $\psi : \mathbb{X} \to \mathbb{R}$ of φ from $\mathbb{U}$ to the whole space $\mathbb{X}$, i. e., there exists a linear functional ψ such that

$$\psi(x) = \varphi(x), \quad \forall x \in \mathbb{U} \quad \textit{and} \quad \psi(x) \le \mathcal{N}(x), \quad \forall x \in \mathbb{X}.$$

The Hahn–Banach separation theorems are the geometrical versions of the Hahn–Banach theorem. The separation theorem is derived from the original form of Theorem 2.3.3.

Let $\mathbb{X}$ be a real vector space, A and B nonempty subsets of $\mathbb{X}$, $f \neq 0$ a real linear functional on $\mathbb{X}$, $s \in \mathbb{R}$ a scalar, and let $V = f^{-1}(\{s\}) := \{x \in \mathbb{X} \mid f(x) = s\}$ be a *hyperplane*. We also define the lower (resp., upper) half-space to be $\{x \in \mathbb{X} \mid f(x) \leq s\}$ (resp., $\{x \in \mathbb{X} \mid f(x) \geq s\}$). We define the strict lower (resp., strict upper) half-space to be $\{x \in \mathbb{X} \mid f(x) < s\}$ (resp., $\{x \in X \mid f(x) > s\}$).

We say that the hyperplane V (or f) separates A and B if $\sup f(A) \leq s \leq \inf f(B)$ or equivalently, if $f(a) \leq s \leq f(b)$ for all $a \in A$ and $b \in B$.

Note that A and B are separated if and only if $\{0\}$ and $B\backslash A$ are separated.

If $a_0 \in A$ and V separates A and $\{a_0\}$, then V is called a supporting hyperplane of A at a_0, a_0 is called a support point of A, and f is called a *support functional*. If A is convex and $a_0 \in A$, then we call a_0 a smooth point of A if there exists a unique hyperplane V such that $a_0 \in A \cap V$. We call a normed space $\mathbb{X}$ smooth if at each point x in its unit ball there exists a unique closed hyperplane to the unit ball at x.

The Hahn–Banach separation theorem is stated below without proof.

Theorem 2.3.4 (Hahn–Banach separation theorem). *Let A and B be nonempty and convex subsets of a real normed vector space $\mathbb{X}$. Furthermore, assume that A and B are disjoint and that A has an interior point. Then there is a hyperplane that separates A and B.*

2.3.1 Bounded linear functionals on $\mathbb{H}$

Recall that $\mathfrak{B}(\mathbb{H})$ and $\mathfrak{T}(\mathbb{H})$ represent, respectively, the space of bounded linear operators (under the operator norm $\|\cdot\|_\infty$) and the space of trace- class operators (under the trace norm $\|\cdot\|_1$) on the complex Hilbert space $\mathbb{H}$.

A bounded linear functional $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathbb{C}$ is said to be real if

$$\Upsilon(\mathbf{a}^*) = \overline{\Upsilon(\mathbf{a})}, \quad \forall \mathbf{a} \in \mathfrak{B}(\mathbb{H}),$$

where $\overline{\Upsilon(\mathbf{a})}$ is the complex conjugate of $\Upsilon(\mathbf{a})$. Υ is said to be positive if

$$\Upsilon(\mathbf{b}^*\mathbf{b}) \geq 0, \quad \forall \mathbf{b} \in \mathfrak{B}(\mathbb{H}),$$

and has mass 1 if $\Upsilon(\mathbf{I}_{\mathbb{H}}) = 1$, where $\mathbf{I}_{\mathbb{H}} \in \mathfrak{B}(\mathbb{H})$ denotes the identity operator on $\mathbb{H}$, i. e., $\mathbf{I}_{\mathbb{H}}\phi = \phi$ for all $\phi \in \mathbb{H}$. A bounded linear functional $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathbb{C}$ is said to be self-adjoint if $\Upsilon = \Upsilon^*$, where Υ^* denotes the adjoint of Υ as defined in Chapter 1. It is known that every linear functional on $\mathfrak{B}(\mathbb{H})$ can be uniquely represented in the form $\Upsilon = \Upsilon_1 + \imath\Upsilon_2$ ($\imath = \sqrt{-1}$), where Υ_1, Υ_2 are real self-adjoint linear functionals. In fact, we may simply let

$$\Upsilon_1 = \frac{\Upsilon + \Upsilon^*}{2} \quad \text{and} \quad \Upsilon_2 = \frac{\Upsilon - \Upsilon^*}{2\imath}.$$

Note that if Υ is positive (denoted by $\Upsilon \geq 0$), it is automatically self-adjoint.

Lemma 2.3.5. *For each $\varphi, \psi \in \mathbb{H}$, define $\mathbf{S}_{\varphi,\psi} : \mathbb{H} \to \mathbb{H}$ by*

$$\mathbf{S}_{\varphi,\psi}\phi = \langle \psi, \phi \rangle_{\mathbb{H}} \varphi := |\varphi\rangle_{\mathbb{H}} \langle \psi | \phi \rangle_{\mathbb{H}}, \quad \forall \phi \in \mathbb{H}. \tag{2.3}$$

Then $\mathbf{S}_{\varphi,\psi}$ is a rank-one operator on $\mathbb{H}$. Furthermore, $\mathbf{S}^*_{\varphi,\psi} = \mathbf{S}_{\psi,\varphi}$, where $\mathbf{S}^*_{\varphi,\psi}$ is the adjoint of the operator $\mathbf{S}_{\varphi,\psi}$, and $\mathbf{S}_{\varphi,\psi}\mathbf{S}_{\psi,\varphi} = \|\psi\|^2_{\mathbb{H}} \mathbf{S}_{\varphi,\varphi}$.

Proof. It is clear that $\mathbf{S}_{\varphi,\psi}$ is a rank-one projection from $\mathbb{H}$ to the one-dimensional subspace $\mathbb{C}\varphi := \{c\varphi \mid c \in \mathbb{C}\}$ of $\mathbb{H}$.

To show that $\mathbf{S}^*_{\varphi,\psi} = \mathbf{S}_{\psi,\varphi}$, we let $\psi_1, \psi_2 \in \mathbb{H}$. Then

$$\begin{aligned}
&\langle \mathbf{S}^*_{\varphi,\psi}\psi_1, \psi_2 \rangle_{\mathbb{H}} \\
&\quad = \langle \psi_1, \mathbf{S}_{\varphi,\psi}\psi_2 \rangle_{\mathbb{H}} \quad \text{(by the definition of adjoint operator)} \\
&\quad = \langle \psi_1, \varphi \langle \psi, \psi_2 \rangle_{\mathbb{H}} \rangle_{\mathbb{H}} \quad \text{(by the definition of } \mathbf{S}_{\varphi,\psi}) \\
&\quad = \langle \psi, \psi_2 \rangle_{\mathbb{H}} \langle \psi_1, \varphi \rangle_{\mathbb{H}} \quad \text{(because } \langle \cdot, \cdot \rangle_{\mathbb{H}} \text{ is linear in second argument)} \\
&\quad = \langle \psi_1, \varphi \rangle_{\mathbb{H}} \langle \psi, \psi_2 \rangle_{\mathbb{H}} \quad \text{(because complex numbers are commutative).}
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\langle \mathbf{S}_{\psi,\varphi}\psi_1, \psi_2 \rangle_{\mathbb{H}} \\
&\quad = \langle \psi \langle \varphi, \psi_1 \rangle_{\mathbb{H}}, \psi_2 \rangle_{\mathbb{H}} \quad \text{(by the definition of } \mathbf{S}_{\psi,\varphi}) \\
&\quad = \overline{\langle \varphi, \psi_1 \rangle_{\mathbb{H}}} \langle \psi, \psi_2 \rangle_{\mathbb{H}} \quad \text{(because } \langle \cdot, \cdot \rangle_{\mathbb{H}} \text{ is conjugate linear in first argument)} \\
&\quad = \langle \psi_1, \varphi \rangle_{\mathbb{H}} \langle \psi, \psi_2 \rangle_{\mathbb{H}} \quad \text{(because } \langle \cdot, \cdot \rangle_{\mathbb{H}} \text{ is conjugate symmetric).}
\end{aligned}$$

Therefore,

$$\langle \mathbf{S}^*_{\varphi,\psi}\psi_1, \psi_2 \rangle_{\mathbb{H}} = \langle \mathbf{S}_{\psi,\varphi}\psi_1, \psi_2 \rangle_{\mathbb{H}}, \quad \forall \psi_1, \psi_2 \in \mathbb{H}.$$

This shows that $\mathbf{S}^*_{\varphi,\psi} = \mathbf{S}_{\psi,\varphi}$.

To show that $\mathbf{S}_{\varphi,\psi}\mathbf{S}_{\psi,\varphi} = \|\psi\|^2_{\mathbb{H}} \mathbf{S}_{\varphi,\varphi}$, we let $\phi \in \mathbb{H}$. Then

$$\begin{aligned}
&(\mathbf{S}_{\varphi,\psi}\mathbf{S}_{\psi,\varphi})\phi \\
&\quad = \mathbf{S}_{\varphi,\psi}(\mathbf{S}_{\psi,\varphi}\phi) \quad \text{(by composition of two operators)} \\
&\quad = \mathbf{S}_{\varphi,\psi}(\langle \varphi, \phi \rangle_{\mathbb{H}} \psi) \quad \text{(by definition of } \mathbf{S}_{\psi,\varphi}) \\
&\quad = \langle \varphi, \phi \rangle_{\mathbb{H}} \mathbf{S}_{\varphi,\psi}\psi \quad \text{(since } \mathbf{S}_{\varphi,\psi} \text{ is a linear operator)} \\
&\quad = \langle \varphi, \phi \rangle_{\mathbb{H}} (\langle \psi, \psi \rangle_{\mathbb{H}} \varphi) \quad \text{(by definition of } \mathbf{S}_{\varphi,\psi})
\end{aligned}$$

$$= \|\psi\|_{\mathbb{H}}^2 \varphi \langle \varphi, \phi \rangle_{\mathbb{H}} \quad \text{(because } \langle \psi, \psi \rangle_{\mathbb{H}} = \|\psi\|_{\mathbb{H}}^2\text{)}$$
$$= \|\psi\|_{\mathbb{H}}^2 \mathbf{S}_{\varphi,\varphi}\phi \quad \text{(by definition of } \mathbf{S}_{\varphi,\varphi}\text{)}.$$

Therefore, $\mathbf{S}_{\varphi,\psi}\mathbf{S}_{\psi,\varphi} = \|\psi\|_{\mathbb{H}}^2 \mathbf{S}_{\varphi,\varphi}$. This proves the lemma. □

We consider the special case of one-dimensional projection $\mathbf{P}_\phi(\cdot) = \mathbf{S}_{\phi,\phi}(\cdot)$, which will be used quite often throughout this book. Using Dirac's "bra" and "ket" notation, $\mathbf{P}_\phi$, the one-dimensional projection along ϕ, can be expressed as $|\phi\rangle_{\mathbb{H}}\langle\phi|$. For illustration purposes, we write

$$\mathbf{P}_\phi(\psi) = \langle \phi, \psi \rangle_{\mathbb{H}} \phi = (|\phi\rangle_{\mathbb{H}}\langle\phi|)\psi = |\phi\rangle\langle\phi|\psi\rangle_{\mathbb{H}} = \langle\phi|\psi\rangle_{\mathbb{H}}|\phi\rangle,$$

where $|\phi\rangle = \phi$ is a vector in $\mathbb{H}$ as mentioned earlier and $\langle\phi| : \mathbb{H} \to \mathbb{C}$ is the bounded linear functional on $\mathbb{H}$ defined by

$$\langle\phi|\psi = \langle\phi|\psi\rangle_{\mathbb{H}} := \langle\phi, \psi\rangle_{\mathbb{H}}.$$

The following proposition shows that the space of trace-class operators, $\mathfrak{T}(\mathbb{H})$, under the σ-weak topology, is the *predual* of $\mathfrak{B}(\mathbb{H})$ in the sense that the topological dual of $\mathfrak{T}(\mathbb{H})$ equals to $\mathfrak{B}(\mathbb{H})$. We write $\mathfrak{B}_*(\mathbb{H}) = \mathfrak{T}(\mathbb{H})$. In this case, $(\mathfrak{B}_*(\mathbb{H}))^* = \mathfrak{B}(\mathbb{H})$.

Recall from Remark 2.1 that $(\mathbf{a}_\alpha)_\alpha$ converges σ-weakly to $\mathbf{a}$ if and only if it converges to $\mathbf{a}$ in weak*-topology. The weak* *topology* (or *σ-weak topology*) is the weak topology arising from the operator pairing $(\mathbf{a}, \mathbf{b}) \mapsto \mathrm{tr}[\mathbf{ab}]$ of $\mathfrak{B}(\mathbb{H})$ with the trace-class operators. That is, it is the weakest topology that makes the map $\mathbf{a} \mapsto \mathrm{tr}[\mathbf{ab}]$ continuous for all trace-class operator $\mathbf{b} \in \mathfrak{T}(\mathbb{H})$. The weak* topology is finer than the weak operator topology but the weak and weak* topology coincide on bounded subsets of $\mathbb{H}$.

Proposition 2.3.6.
1. *The space* $\mathfrak{B}(\mathbb{H})$ *is the (topological) dual of* $\mathfrak{T}(\mathbb{H})$.
2. $\mathfrak{T}(\mathbb{H})$ *is the predual of* $\mathfrak{B}(\mathbb{H})$ *(i. e.,* $\mathfrak{T}^*(\mathbb{H}) = \mathfrak{B}(\mathbb{H})$*) by the duality*

$$(\mathbf{a}, \mathbf{T}) \in \mathfrak{B}(\mathbb{H}) \times \mathfrak{T}(\mathbb{H}) \mapsto \ll \mathbf{a}, \mathbf{T} \gg := \mathrm{tr}[\mathbf{aT}].$$

3. *The* weak* *topology on* $\mathfrak{B}(\mathbb{H})$ *arising from this duality coincide with the σ-weak topology on* $\mathfrak{B}(\mathbb{H})$.

Proof. (1) First, we want to show that $\mathfrak{T}^*(\mathbb{H}) := (\mathfrak{T}(\mathbb{H}))^* = \mathfrak{B}(\mathbb{H})$. Let $(\mathbf{a}, \mathbf{T}) \in \mathfrak{B}(\mathbb{H}) \times \mathfrak{T}(\mathbb{H})$. From the inequality,

$$| \ll \mathbf{a}, \mathbf{T} \gg | = |\,\mathrm{tr}[\mathbf{aT}]| \le \|\mathbf{a}\|_\infty \|\mathbf{T}\|_1,$$

it shows that $\ll \mathbf{a}, \cdot \gg : \mathfrak{T}(\mathbb{H}) \to \mathbb{C}$ is a bounded linear functional on $\mathfrak{T}(\mathbb{H})$. Therefore, $\mathbf{a} \in \mathfrak{T}^*(\mathbb{H})$. This shows that $\mathfrak{B}(\mathbb{H})$ is a subspace of $\mathfrak{T}^*(\mathbb{H})$. Conversely, we want to show that $\mathfrak{T}^*(\mathbb{H}) \subset \mathfrak{B}(\mathbb{H})$. Consider the rank one operator $\mathbf{S}_{\varphi,\psi}$ on $\mathbb{H}$ defined by (2.3) for all $\varphi, \psi \in \mathbb{H}$. It follows from Lemma 2.3.5 that $\mathbf{S}^*_{\varphi,\psi} = \mathbf{S}_{\psi,\varphi}$ and $\mathbf{S}_{\varphi,\psi}\mathbf{S}_{\psi,\varphi} = \|\psi\|^2_{\mathbb{H}}\mathbf{S}_{\varphi,\varphi}$. Hence,

$$\begin{aligned}
&\|\mathbf{S}_{\varphi,\psi}\|_1 \\
&\quad = \operatorname{tr}[|\mathbf{S}_{\varphi,\psi}|] \quad (\text{by definition of } \|\cdot\|_1) \\
&\quad = \operatorname{tr}[\sqrt{\mathbf{S}^*_{\varphi,\psi}\mathbf{S}_{\varphi,\psi}}] \quad (\text{by definition of } |\mathbf{S}_{\varphi,\psi}|) \\
&\quad = \operatorname{tr}[\sqrt{\mathbf{S}_{\psi,\varphi}\mathbf{S}_{\varphi,\psi}}] \quad (\text{by Lemma 2.3.5}) \\
&\quad = \operatorname{tr}[\sqrt{\|\varphi\|^2_{\mathbb{H}}\mathbf{S}_{\psi,\psi}}] \quad (\text{by Lemma 2.3.5}) \\
&\quad = \|\varphi\|_{\mathbb{H}}(\operatorname{tr}[\sqrt{\mathbf{S}_{\psi,\psi}}]) \quad (\text{by linearity of the trace operator}) \\
&\quad \le \|\varphi\|_{\mathbb{H}}\left(\sum_{i=1}^{\infty}\langle\psi\langle\psi, e_i\rangle_{\mathbb{H}} e_i\rangle_{\mathbb{H}}\right)^{1/2}, \\
&\qquad (\text{where } (e_i)_{i=1}^{\infty} \text{ is any orthonormal basis of } \mathbb{H}) \\
&\quad = \|\varphi\|_{\mathbb{H}}\left(\sum_{i=1}^{\infty}\langle\psi, e_i\rangle^2_{\mathbb{H}}\right)^{1/2} \quad (\text{by linearity of the second argument in } \langle\cdot,\cdot\rangle_{\mathbb{H}}) \\
&\quad = \|\psi\|_{\mathbb{H}}\|\varphi\|_{\mathbb{H}}.
\end{aligned}$$

It follows that

$$|\omega(\mathbf{S}_{\varphi,\psi})| \le \|\omega\|_\infty \|S_{\varphi,\psi}\|_1 \le \|\omega\|_\infty \|\varphi\|_{\mathbb{H}}\|\psi\|_{\mathbb{H}}, \quad \forall \omega \in \mathfrak{T}^*(\mathbb{H}),$$

where $\|\omega\|_\infty$ denotes the operator norm of ω. This implies that ω is a bounded linear operator on the space of rank-one projection operators $\{\mathbf{S}_{\varphi,\phi} \mid \varphi, \phi \in \mathbb{H}\}$. Hence, there exists, by the Riesz representation theorem (Theorem 2.3.1), an $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$ with $\|\mathbf{a}\|_\infty \le \|\omega\|_\infty$ such that

$$\omega(\mathbf{S}_{\varphi,\psi}) = \langle\psi, \mathbf{a}\varphi\rangle_{\mathbb{H}}.$$

Consider $\omega_0 \in \mathfrak{T}^*(\mathbb{H})$ defined by

$$\omega_0(\mathbf{T}) = \ll \mathbf{a}, \mathbf{T} \gg = \operatorname{tr}[\mathbf{aT}];$$

then

$$\begin{aligned}
\omega_0(\mathbf{S}_{\varphi,\psi}) &= \operatorname{tr}[\mathbf{a}\mathbf{S}_{\varphi,\psi}] \\
&= \langle\psi, \mathbf{a}\varphi\rangle_{\mathbb{H}} \\
&= \omega(\mathbf{S}_{\varphi,\psi}).
\end{aligned}$$

Now for any $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$ there exist bounded sequences $(\psi_n)_n$ and $(\varphi_n)_n$ and a sequence of complex numbers such that

$$\sum_n |\alpha_n| < \infty$$

and

$$\mathbf{T} = \sum_n \alpha_n \mathbf{S}_{\varphi_n,\psi_n}.$$

The latter series converges with respect to the trace norm, and hence,

$$\omega(\mathbf{T}) = \sum_n \alpha_n \omega(\mathbf{S}_{\varphi_n,\psi_n}) = \sum_n \alpha_n \omega_0(\mathbf{S}_{\varphi_n,\psi_n}) = \omega_0(\mathbf{T}) = \mathrm{tr}[\mathbf{aT}].$$

Thus, $\mathfrak{B}(\mathbb{H})$ is just the dual of $\mathfrak{T}(\mathbb{H})$.

(3). The weak* topology on $\mathfrak{B}(\mathbb{H})$ arising from this duality is given by the seminorms

$$\mathbf{a} \in \mathfrak{B}(\mathbb{H}) \mapsto |\,\mathrm{tr}[\mathbf{aT}]|.$$

Now for

$$\mathbf{T} = \sum_n \alpha_n \mathbf{S}_{\varphi_n,\psi_n}$$

one has

$$\mathrm{tr}[\mathbf{aT}] = \sum_n \alpha_n \,\mathrm{tr}[\mathbf{S}_{\varphi_n,\psi_n}\mathbf{a}] = \sum_n \alpha_n \langle \psi_n, \mathbf{a}\varphi_n\rangle_{\mathbb{H}}.$$

Thus, the seminorms are equivalent to the seminorms defining the σ-weak topology. This proves the proposition. □

We have the following easy consequence of Proposition 2.3.6.

Corollary 2.3.7. *The space of trace-class operators* $\mathfrak{T}(\mathbb{H})$ *is a complex Banach space under trace norm* $\|\cdot\|_1$.

2.3.2 $\mathcal{A}^*$ and $\mathcal{A}_*$

Let $\mathcal{A} \subseteq \mathfrak{B}(\mathbb{H})$ be either a C^*-algebra or a von Neumann algebra of bounded linear operators on $\mathbb{H}$.

We define normal and σ-weakly continuous linear functionals below.

Definition 2.3.8. A positive linear functional $\rho : \mathcal{A} \subseteq \mathfrak{B}(\mathbb{H}) \to \mathbb{C}$ is said to be normal if $\rho(\vee_\alpha \mathbf{a}_\alpha) = \vee_\alpha \rho(\mathbf{a}_\alpha)$ for any upper-bounded increasing net $\{\mathbf{a}_\alpha\}_{\alpha\in L}$ of positive elements in $\mathcal{A} \subseteq \mathfrak{B}(\mathbb{H})$, where $\vee_\alpha(\cdot)$ denotes the least upper bound of the net (see Definition 2.1.1 and Zorm's lemma 2.1.3).

Definition 2.3.9. Let ρ be a positive linear functional on $\mathcal{A} \subseteq \mathfrak{B}(\mathbb{H})$. We say that ρ is σ-weakly continuous if for every increasing net of $(\mathbf{a}_\alpha) \subset \mathcal{A}$ that converges to $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$ in σ-weakly topology (see Definition 2.1.2 for the definition of σ-weakly convergence), we have

$$\lim_\alpha \rho(\mathbf{a}_\alpha) = \rho(\mathbf{a}).$$

It is well-known fact (see (2) of Remark 2.6) that a net $(\mathbf{a}_\alpha)_\alpha$ in $\mathcal{A}$ is σ-weakly convergent to $\mathbf{a} \in \mathcal{A}$ if and only if $\operatorname{tr}[\rho(\mathbf{a}_\alpha - \mathbf{a})] \to 0$ for all $\rho \in \mathfrak{T}(\mathbb{H})$.

We state the following useful *Beppo Levi theorem* (also better known as monotone convergence theorem) below without proof.

Theorem 2.3.10 (Beppo Levi theorem). *Let $(\mathbb{X}, \Sigma, \mu)$ be a measure space. Let $(f_n)_{n=1}^{+\infty}$ be an increasing sequence of positive Σ-measurable functions, $f_n : \mathbb{X} \to [0, \infty]$ for $n \geq 1$. Then*

$$\int_{\mathbb{X}} \Big(\sup_{n\geq 1} f_n(x)\Big)\mu(dx) = \sup_{n\geq 1}\Big(\int_{\mathbb{X}} f_n(x)\mu(dx)\Big).$$

The following equivalent fundamental properties of states on a von Neumann algebra are well-known (see Bratteli and Robinson [15] Theorem 2.4.21 on p. 76 and also Chang [24]).

Theorem 2.3.11. *Let $\mathcal{A}$ be a von Neumann algebra of bounded linear operators on a Hilbert space $\mathbb{H}$ and let ω be a bounded linear functional on $\mathcal{A}$. The following conditions are equivalent:*

1. *ω is normal, i. e., $\omega(\vee_\alpha \mathbf{a}_\alpha) = \vee_\alpha \omega(\mathbf{a}_\alpha)$ for any upper-bounded increasing net of bounded operators $(\mathbf{a}_\alpha)_{\alpha\in L} \subset \mathfrak{B}(\mathbb{H})$, where $\vee_\alpha \mathbf{a}_\alpha := \sup_\alpha \mathbf{a}_\alpha$;*
2. *ω is σ-weakly continuous, i. e., $\lim_\alpha \omega(\mathbf{a}_\alpha) = \omega(\mathbf{a})$ for any net $(\mathbf{a}_\alpha)_{\alpha\in L}$ that converges to $\mathbf{a}$ in σ-weak topology;*
3. *there exists a density operator ρ (i. e., a positive trace-class operator on $\mathbb{H}$ with $\operatorname{tr}[\rho] = 1$) such that $\omega(\mathbf{a}) = \operatorname{tr}[\rho\mathbf{a}]$ for all $\mathbf{a} \in \mathcal{A}$.*
4. *$\omega(\sum_{i\in I} \mathbf{p}_i) = \sum_{i\in I} \omega(\mathbf{p}_i)$ for every family $\{\mathbf{p}_i, i \in I\}$ of pairwise orthogonal projections in $\mathcal{A}$.*

Proof. (3) $\Rightarrow$ (2). Suppose there exists a positive $\rho \in \mathfrak{T}(\mathbb{H})$ with $\operatorname{tr}[\rho] = 1$ such that $\omega(\mathbf{a}) := \operatorname{tr}[\rho\mathbf{a}]$ for all $\mathbf{a} \in \mathcal{A}$. Then the functional $\omega : \mathcal{A} \to \mathbb{C}$ defined above is continuous under the weak* topology. This shows that the functional ω is continuous under the σ-weak topology by Proposition 2.3.6.

(2) ⇒ (1) follows from Proposition 2.1.4.

(2) ⇒ (3). If ω is σ-weakly continuous, there exist sequences $(\zeta_n)_{n=1}^{+\infty}, (\xi_n)_{n=1}^{+\infty}$ of vectors in $\mathbb{H}$ such that $\sum_{n=1}^{+\infty} \|\zeta_n\|_{\mathbb{H}}^2 < +\infty$, $\sum_{n=1}^{+\infty} \|\zeta_n\|_{\mathbb{H}}^2 < +\infty$ and $\omega(\mathbf{a}) = \sum_{n=1}^{+\infty} \langle \zeta_n, \mathbf{a}\xi_n \rangle_{\mathbb{H}}$. Define $\tilde{\mathbb{H}} = \bigoplus_{n=1}^{+\infty} \mathbb{H}_n$ (see (2.8) for the definition of direct sum of Hilbert spaces), where $\mathbb{H}_n = \mathbb{H}$ for all $n \geq 1$, and introduce a representation π of $\mathcal{A}$ on $\tilde{\mathbb{H}}$ by $\pi(\mathbf{a})(\bigoplus_{n=1}^{+\infty} \psi_n) = \bigoplus_{n=1}^{+\infty}(\mathbf{a}\psi_n)$. Let $\zeta = \bigoplus_{n=1}^{+\infty} \zeta_n$, $\eta = \bigoplus_{n=1}^{+\infty} \eta_n$ and then $\omega(\mathbf{a}) = \langle \zeta, \pi(\mathbf{a})\eta \rangle_{\mathbb{H}}$. Since $\omega(\mathbf{a})$ is real for $\mathbf{a} \in \mathcal{A}_+$, we have

$$\begin{aligned} 4\omega(\mathbf{a}) &= 2\langle \zeta, \pi(\mathbf{a})\eta \rangle_{\mathbb{H}} + 2\langle \zeta, \pi(\mathbf{a}^*)\eta \rangle_{\mathbb{H}} \\ &= 2\langle \zeta, \pi(\mathbf{a})\eta \rangle_{\mathbb{H}} + 2\langle \eta, \pi(\mathbf{a})\zeta \rangle_{\mathbb{H}} \\ &= \langle \zeta + \eta, \pi(\mathbf{a})(\zeta + \eta) \rangle_{\mathbb{H}} - \langle \zeta - \eta, \pi(\mathbf{a})(\zeta - \eta) \rangle_{\mathbb{H}} \\ &\leq \langle \zeta + \eta, \pi(\mathbf{a})(\zeta + \eta) \rangle_{\mathbb{H}} \end{aligned}$$

Hence, by Proposition 2.5.2 there exists a positive $\mathbf{T} \in \pi'(\mathcal{A})$ with $0 \leq \mathbf{T} \leq \mathbf{I}/2$ such that

$$\begin{aligned} \langle \zeta, \pi(\mathbf{a})\eta \rangle_{\mathbb{H}} &= \langle \mathbf{T}(\zeta + \eta), \pi(\mathbf{a})\mathbf{T}(\zeta + \eta) \rangle_{\mathbb{H}} \\ &= \langle \psi, \pi(\mathbf{a})\psi \rangle_{\mathbb{H}}. \end{aligned}$$

The right-hand side of this relation can be used to extend ω to a σ-weakly continuous positive linear functional $\tilde{\omega}$ on $\mathfrak{B}(\mathbb{H})$. Since $\tilde{\omega}(\mathbf{I}_{\mathbb{H}}) = 1$, by Proposition 2.3.6 there exists a trace-class operator ρ with $\mathrm{tr}[\rho] = 1$ such that

$$\tilde{\omega}(\mathbf{a}) = \mathrm{tr}[\rho\mathbf{a}].$$

Let $\mathbf{p}$ be the rank one projection operator with range ζ. Then

$$\langle \zeta, \rho\zeta \rangle_{\mathbb{H}} = \mathrm{tr}[\mathbf{p}\rho\mathbf{p}] = \mathrm{tr}[\rho\mathbf{p}] = \tilde{\omega}(\mathbf{p}) \geq 0.$$

Thus, ρ is positive.

(1) ⇒ (2). Assume that the positive bounded linear functional ω is normal on $\mathcal{A}$. Let $(\mathbf{b}_\alpha)$ be an increasing net of elements in $\mathcal{A}_+$ such that $\|\mathbf{b}_\alpha\|_\infty \leq 1$ for all α and such that $\mathbf{a} \mapsto \omega(\mathbf{a}\mathbf{b}_\alpha)$ is σ-strongly continuous for all α. We can use Proposition 2.3.6 to define $\mathbf{b}$ by

$$\mathbf{b} = \vee_\alpha \mathbf{b}_\alpha = (\sigma\text{-strong}) \lim_\alpha \mathbf{b}_\alpha.$$

Then $0 \leq \mathbf{b} \leq \mathbf{I}$ and $\mathbf{b} \in \mathcal{A}$. But for all $\mathbf{a} \in \mathcal{A}$, we have

$$\begin{aligned} |\omega(\mathbf{a}\mathbf{b} - \mathbf{a}\mathbf{b}_\alpha)|^2 &= |\omega(\mathbf{a}(\mathbf{b} - \mathbf{b}_\alpha)^{1/2}(\mathbf{b} - \mathbf{b}_\alpha)^{1/2})| \\ &\leq \omega(\mathbf{a}(\mathbf{b} - \mathbf{b}_\alpha)\mathbf{a}^*)\omega(\mathbf{b} - \mathbf{b}_\alpha) \\ &\leq \|\mathbf{a}\|_\infty^2 \omega(\mathbf{b} - \mathbf{b}_\alpha). \end{aligned}$$

Hence,

$$\|\omega(\cdot\mathbf{b}) - \omega(\cdot\mathbf{b}_\alpha)\|_\infty := \sup_{\|\mathbf{a}\|=1} |\omega(\mathbf{ab} - \mathbf{ab}_\alpha)| \le \big(\omega(\mathbf{b} - \mathbf{b}_\alpha)\big)^{1/2}.$$

Since ω is normal, $\omega(\mathbf{b} - \mathbf{b}_\alpha) \to 0$ and $\omega(\cdot\mathbf{b}_\alpha)$ tends to $\omega(\cdot\mathbf{b})$ in $\|\cdot\|_\infty$-norm. As $\mathcal{A}_*$ is a Banach space, $\omega(\cdot\mathbf{b}_\alpha) \in \mathcal{A}_*$. Now, applying Zorn's lemma (see Lemma 2.1.3, we can find a maximal element $\mathbf{P} \in \mathcal{A}_+ \cap \{\mathbf{a} \in \mathcal{A} \mid \|\mathbf{a}\|_\infty = 1\}$ such that $\mathbf{a} \mapsto \omega(\mathbf{aP})$ is σ-strong continuous. We consider the following two cases: (i) $\mathbf{P} = \mathbf{I}$ and (ii) $\mathbf{P} \neq \mathbf{I}$. If (i) $\mathbf{P} = \mathbf{I}$, then the theorem is proved. If (ii) $\mathbf{P} \neq \mathbf{I}$, we put $\mathbf{P}' = \mathbf{I} - \mathbf{P}$ and choose an $\zeta \in \mathbb{H}$ such that $\omega(\mathbf{P}') < \langle\zeta, \mathbf{P}'\zeta\rangle_{\mathbb{H}}$. If $(\mathbf{b}_\alpha)$ be an increasing net of elements in $\mathcal{A}_+$ such that $\mathbf{b}_\alpha \le \mathbf{P}'$, $\omega(\mathbf{B}_\alpha) \ge \langle\zeta, \mathbf{b}_\alpha\zeta\rangle_{\mathbb{H}}$ and

$$\mathbf{b} = \vee_\alpha \mathbf{b}_\alpha = (\sigma\text{-strong}) \lim_\alpha \mathbf{b}_\alpha,$$

then $\mathbf{b} \in \mathcal{A}_+$, $\mathbf{b} \le \mathbf{P}'$ and $\omega(\mathbf{b}) = \sup \omega(\mathbf{b}_\alpha) \ge \sup\langle\zeta, \mathbf{b}\zeta\rangle_{\mathbb{H}}$. Hence, by Zorn's lemma (see Lemma 2.1.3), there exists a maximal $\mathbf{b} \in \mathcal{A}_+$ such that $\mathbf{b} \le \mathbf{P}'$ and $\omega(\mathbf{b}) \ge \langle\zeta, \mathbf{b}\zeta\rangle_{\mathbb{H}}$. Put $\mathbf{Q} = \mathbf{P}' - \mathbf{b}$. Then $\mathbf{Q} \in \mathcal{A}_+$, $\mathbf{Q} \neq \mathbf{0}$ (since $\omega(\mathbf{P}') < \langle\zeta, \mathbf{P}'\zeta\rangle_{\mathbb{H}}$), and if $\mathbf{a} \in \mathcal{A}_+$, $\mathbf{a} \le \mathbf{Q}$, $\mathbf{a} \neq \mathbf{0}$, then $\omega(\mathbf{a}) < \langle\zeta, \mathbf{a}\zeta\rangle_{\mathbb{H}}$ by the maximality of $\mathbf{b}$.

For any $\mathbf{a} \in \mathcal{A}$, we have

$$\mathbf{Qa}^*\mathbf{aQ} \le \|\mathbf{a}\|_\infty^2 \mathbf{Q}^2 \le \|\mathbf{a}\|_\infty^2 \|\mathbf{Q}\|_\infty \mathbf{Q}.$$

Hence,

$$\frac{(\mathbf{Qa}^*\mathbf{aQ})}{\|\mathbf{a}\|_\infty^2\|\mathbf{Q}\|_\infty} \le \mathbf{Q} \quad \text{and} \quad \omega(\mathbf{Qa}^*\mathbf{aQ}) < \langle\zeta, \mathbf{Qa}^*\mathbf{aQ}\zeta\rangle_{\mathbb{H}}.$$

Combining this with the Cauchy–Schwarz inequality (see (1.2)), one finds

$$|\omega(\mathbf{aQ})|^2 \le \omega(\mathbf{1})\omega(\mathbf{Qa}^*\mathbf{aQ}) < \langle\zeta, \mathbf{Qa}^*\mathbf{aQ}\zeta\rangle_{\mathbb{H}} = \|\mathbf{aQ}\zeta\|_{\mathbb{H}}^2.$$

Thus, both $\mathbf{a} \mapsto \omega(\mathbf{aQ})$ and $\mathbf{a} \mapsto \omega(\mathbf{a}(\mathbf{P} + \mathbf{Q}))$ are σ-strong continuous, since $\mathbf{P} + \mathbf{Q} \le \mathbf{1}$. This contradicts the maximality of $\mathbf{P}$. Therefore, $\mathbf{P} = \mathbf{I}$.

(1) $\Leftrightarrow$ (4). This follows easily from the definition of the normality of ω and the Beppo Levy Theorem 2.3.10) This proves the theorem. □

We recall that a bounded linear operator $\mathbf{T} : \mathbb{H} \to \mathbb{H}$ is said to be diagonalizable if there exists an orthonormal basis $(e_i)_{i=1}^{+\infty}$ consisting of eigenvectors of $\mathbf{T}$. An equivalent definition of diagonalizable operator $\mathbf{T}$ is that there exists an orthonormal basis $(e_i)_{i=1}^{+\infty}$ of $\mathbb{H}$ and a sequence of complex numbers $(\lambda_i)_{i=1}^{+\infty} \subset \mathbb{C}$ such that

$$\mathbf{T}(\phi) = \sum_{i=1}^{+\infty} \lambda_i \langle\phi, e_i\rangle_{\mathbb{H}} e_i, \quad \forall\phi \in \mathbb{H}.$$

Corollary 2.3.12. *If ω is a normal bounded linear functional on $\mathcal{A}$, then there exists an orthonormal part $(e_n)_{n=0}^{+\infty}$ of $\mathbb{H}$ and a sequence $(\lambda_n)_{n=0}^{+\infty}$ such that* $\lim_{n\to+\infty}\lambda_n = 0$, $\sum_{n=0}^{\infty}\lambda_n = 1$ *and*

$$\omega(\mathbf{a}) = \operatorname{tr}\Big[\sum_n \lambda_n |e_n\rangle_{\mathbb{H}}\langle e_n|\mathbf{a}\Big], \quad \forall \mathbf{a} \in \mathcal{A}.$$

Proof. This is a trivial consequence of Theorem 2.3.11 since any trace-class operator is compact and any normal compact operator is diagonalizable (see, e. g., Reed and Simon [128]). □

Let $\mathcal{A}_*$ be the predual of the von Neumann algebra $\mathcal{A}$ of bounded linear operators on the complex Hilbert space $\mathbb{H}$, and let $(\mathcal{A}_*)_+$ be the collection positive operators in $\mathcal{A}_*$.

For the following theorem, I denotes a linearly ordered net under relation "$\succ$," a net $(\omega_i)_{i\in I}$ is said to be a decreasing net if $\omega_i \geq \omega_j$ whenever $j \succ i$.

Theorem 2.3.13. *If $(\omega_i)_{i\in I}$ is a decreasing net in $(\mathcal{A}_*)_+$, then $(\omega_i)_{i\in I}$ is weakly convergent in $(\mathcal{A}_*)_+$.*

Proof. Notice that, for $\mathbf{x} \in \mathcal{A}_+$, $(\omega_i(\mathbf{x}))_{i\in I}$ is a decreasing and positive net of real numbers and, therefore, is convergent. Therefore, given $\mathbf{a} \in \mathcal{A}$, $\mathbf{a} = \mathbf{b} + \imath\mathbf{c}$ (where $\imath = \sqrt{-1}$) with $\mathbf{b}, \mathbf{c} \in \mathcal{A}_{sa}$, $\mathbf{b} = \mathbf{b}_+ - \mathbf{b}_-$ and $\mathbf{c} = \mathbf{c}_+ - \mathbf{c}_-$, $\mathbf{b}_+, \mathbf{b}_-, \mathbf{c}_+, \mathbf{c}_- \in \mathcal{A}_+$, we can define

$$\omega(\mathbf{a}) := \lim_i\big(\omega_i(\mathbf{b}_+) - \omega_i(\mathbf{b}_-)\big) + \imath \lim_i\big(\omega_i(\mathbf{c}_+) - \omega_i(\mathbf{c}_-)\big) = \lim_i \omega_i(\mathbf{a}).$$

We have to show that ω is normal. By (4) of Theorem 2.3.11, it is sufficient to prove that $\omega(\sum_j \mathbf{p}_j) = \sum_j \omega(\mathbf{p}_j)$, where $(\mathbf{p}_j)_j$ is a sequence of pairwise orthogonal projections. But

$$\omega\Big(\sum_j \mathbf{p}_j\Big) = \lim_i \omega_i\Big(\sum_j \mathbf{p}_j\Big) = \lim_i \sum_j \omega_i(\mathbf{p}_j) = \sum_j \lim_i \omega_i(\mathbf{p}_j) = \sum_j \omega(\mathbf{p}_j)$$

by the Beppo Levi Theorem 2.3.10. This proves the theorem. □

We comment here that the weak*- topology on $\mathcal{A}$ is nothing but the relative σ-weak topology on $\mathcal{A}$. Indeed, since $\mathcal{A}$ is σ-weakly closed, it is enough to show that a net $(\mathbf{a}_\alpha)_\alpha$ in $\mathcal{A}$ converges σ-weakly to $\mathbf{a} \in \mathcal{A}$ if and only if $w^*\text{-}\lim_\alpha \theta(\mathbf{a}_\alpha) = \theta(\mathbf{a})$ on $(\mathcal{A}_*)^*$, where $\theta(\mathbf{b})(\omega) = \omega(\mathbf{b})$ for all $\omega \in \mathcal{A}_*$, $\mathbf{b} \in \mathcal{A}$. But $w^*\text{-}\lim_\alpha \theta(\mathbf{a}_\alpha) = \theta(\mathbf{a})$ on $(\mathcal{A}_*)^*$ if and only if $\omega(\mathbf{a}_\alpha) \to \omega(\mathbf{a})$ for all $\omega \in \mathcal{A}_*$ if and only if $\operatorname{tr}[\rho\mathbf{a}_\alpha] \to \operatorname{tr}[\rho\mathbf{a}]$ for all $\rho \in \mathcal{L}^1(\mathbb{H})$ if and only if $\mathbf{a}_\alpha \to \mathbf{a}$, σ-weakly.

We shall often use a consequence of the equivalence of (1) and (2) in Theorem 2.3.11.

If $\mathcal{A} \subseteq \mathfrak{B}(\mathbb{H})$ is a von Neumann algebra, the subspace $\mathcal{A}_*$ of all σ-weakly continuous functionals in $\mathfrak{T}(\mathbb{H})$ is said to be the predual of $\mathcal{A}$. Since any σ-weakly continuous functional clearly is also norm continuous, $\mathcal{A}_*$ is the subspace of $\mathcal{A}^*$, where $\mathcal{A}^*$ is the space of bounded linear functionals on $\mathcal{A}$ under the operator norm. It follows by Theorem 2.3.11, if

$$\mathcal{A}^{\perp} := \{\rho \in \mathfrak{T}(\mathbb{H}) \mid \mathrm{tr}[\rho \mathbf{a}] = 0 \; \forall \mathbf{a} \in \mathcal{A}\},$$

we have a well-defined linear bijection $\mathfrak{T}(\mathbb{H})/\mathcal{A}^{\perp} \to \mathcal{A}_*$ defined by $\rho + \mathcal{A}^{\perp} \mapsto \omega$, where $\omega(\mathbf{a}) = \mathrm{tr}[\rho\mathbf{a}]$ for all $\mathbf{a} \in \mathcal{A}$. It is easy to check that this map is isometric. Therefore, $\mathcal{A}_*$ can be identified with the Banach space $\mathfrak{T}(\mathbb{H})/\mathcal{A}^{\perp}$ endowed with the quotient norm. Therefore, we have the following.

Proposition 2.3.14. *The predual $\mathcal{A}_*$ of $\mathcal{A}$ is a Banach space in the norm of $\mathcal{A}^*$, and $\mathcal{A}$ is the dual of $\mathcal{A}_*$ with respect to the pairing $\mathcal{A} \times \mathcal{A}_*$ by defining* $(\mathbf{a}, \omega) \mapsto \omega(\mathbf{a}) \in \mathbb{C}$.

Proof. For $\mathbf{a} \in \mathcal{A}$, let

$$\|\mathbf{a}\|_{\mathrm{dual}} = \sup_{\omega \in \mathcal{A}_* : \|\omega\|_1 = 1} |\omega(\mathbf{a})|$$

denote the norm of $\mathbf{a}$ for the duality expressed in the statement of the proposition. Then $\|\mathbf{a}\|_{\mathrm{dual}} \le \|\mathbf{a}\|_{\infty}$. On the other hand, if we denote by $\omega_{u,v}$ the linear functional $\mathfrak{B}(\mathbb{H})$ given by $\omega_{u,v}(\mathbf{a}) = \langle v, \mathbf{a}u\rangle_{\mathbb{H}}$, $u, v \in \mathbb{H}$, we have

$$\|\mathbf{a}\|_{\infty} = \sup_{\|u\|_{\mathbb{H}} = \|v\|_{\mathbb{H}} = 1} \left|\langle v, \mathbf{a}u\rangle_{\mathbb{H}}\right| \le \|\mathbf{a}\|_{\mathrm{dual}},$$

since the restriction of $\omega_{u,v}$ to $\mathcal{A}$ is σ-weakly continuous. Thus, $\mathcal{A}$ can be identified isometrically to a linear subspace of $(\mathcal{A}_*)^*$. We simply have to prove that this identification is onto. Let Ψ be a norm continuous functional on $\mathcal{A}_*$. Since $\mathcal{A}_* \subseteq \mathfrak{B}_*(\mathbb{H}) = \mathfrak{T}(\mathbb{H})$, by the Hahn–Banach theorem (Theorem 2.3.3) we can extend Ψ to a norm continuous functional on $\mathfrak{B}_*(\mathbb{H})$. Therefore, there exists $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$ such that $\Psi(\omega) = \omega(\mathbf{a})$ for all $\omega \in \mathfrak{B}_*(\mathbb{H})$. Therefore, if $\mathbf{x}$ is a self-adjoint element of $\mathcal{A}'$ ($\mathcal{A}'$ is the commutant of $\mathcal{A}$), we have in particular

$$\langle v, \mathbf{a}\mathbf{x}u\rangle_{\mathbb{H}} = \Psi(\omega_{\mathbf{x}u,v}|_{\mathcal{A}}) = \Psi(\omega_{u,\mathbf{x}v}|_{\mathcal{A}}) = \langle \mathbf{x}v, \mathbf{a}u\rangle_{\mathbb{H}} = \langle v, \mathbf{x}\mathbf{a}u\rangle_{\mathbb{H}}$$

for all $u, v \in \mathbb{H}$. Therefore, we have $\mathbf{a} \in \mathcal{A}'' = \mathcal{A}$. □

Interested readers are referred to Chang [24] for further reading on $\mathcal{A}$ and $\mathcal{A}_*$.

2.4 Quantum states

In the following, an important ingredient in quantum communication is a *quantum state*, which summarizes the status of a physical system $\mathbb{H}$ and permits the calculation of statistical quantities (such as probabilities, expectations, correlations) of observables. This section addresses the formulation of von Neumann's postulate 2.

In the following, we assume that the C^*-algebra $\mathcal{A} = \mathfrak{B}(\mathbb{H})$. In this case, $\mathcal{A}_* = \mathfrak{T}(\mathbb{H})$, where $\mathfrak{T}(\mathbb{H})$ denotes the Banach space of trace-class operators under the trace-class norm $\|\cdot\|_1$.

The mathematical definition of a quantum state of a quantum system $\mathbb{H}$ is given below.

Definition 2.4.1. A (quantum) state on a complex Hilbert space $\mathbb{H}$ is a positive (and hence self-adjoint) trace-class operator ρ on $\mathbb{H}$ that has unit-trace, i. e., $\mathrm{tr}[\rho] = 1$.

The set of quantum states on $\mathbb{H}$ will be denoted by $\mathcal{S}(\mathbb{H})$. That is,

$$\mathcal{S}(\mathbb{H}) = \{\rho \in \mathfrak{T}(\mathbb{H}) \mid \rho \geq 0, \mathrm{tr}[\rho] = 1\}. \tag{2.4}$$

We have the following property (see Chang [23]) regarding the set of quantum states $\mathcal{S}(\mathbb{H})$.

Proposition 2.4.2. *Let $\mathbb{H}$ be a complex separable Hilbert space. Then:*
1. *The set of quantum states $\mathcal{S}(\mathbb{H})$ on $\mathbb{H}$ is a compact convex subset of the real vector space $\mathfrak{T}(\mathbb{H})$.*
2. *If $\phi \in \mathbb{H}$ is a unit vector (i. e., $\|\phi\|_{\mathbb{H}} = 1$), then the one-dimensional projection (along the vector ϕ),*

$$\mathbf{P}_\phi := |\phi\rangle_{\mathbb{H}}\langle\phi| : \mathbb{H} \to \mathbb{H},$$

is a quantum state.

Proof. (1) First, we prove that $\mathcal{S}(\mathbb{H})$ is a convex subset of $\mathfrak{T}(\mathbb{H})$. Let $\rho, \omega \in \mathcal{S}(\mathbb{H})$. That is, $\rho, \omega : \mathbb{H} \to \mathbb{C}$ are bounded linear functionals such that $\mathrm{tr}[\rho] = \mathrm{tr}[\omega] = 1$. For any $0 \leq a \leq 1$, it is clear that $a\rho + (1-a)\omega : \mathbb{H} \to \mathbb{C}$ is a bounded linear functional. Furthermore,

$$\mathrm{tr}[a\rho + (1-a)\omega] = \mathrm{tr}[a\rho] + [(1-a)\omega] = a\,\mathrm{tr}[\rho] + (1-a)\,\mathrm{tr}[\omega] = 1.$$

This shows that $a\rho + (1-a)\omega \in \mathcal{S}(\mathbb{H})$. Therefore, $\mathcal{S}(\mathbb{H})$ is convex. Now $\mathcal{S}(\mathbb{H}) \subset \mathfrak{T}(\mathbb{H}) = \mathfrak{B}_*(\mathbb{H})$ is a closed unit ball (because $\|\rho\|_1 = \mathrm{tr}[\rho] = 1$ for all $\rho \in \mathcal{S}(\mathbb{H})$). Therefore, $\mathcal{S}(\mathbb{H})$ is weakly compact in the space $\mathfrak{T}(\mathbb{H})$ by the Banach–Alaoglu theorem (see Theorem 1.1.4).

(2). It is clear that any one-dimensional projection is a positive trace-class operator. Therefore, to show that the one-dimensional projection $\mathbf{P}_\phi := |\phi\rangle_{\mathbb{H}}\langle\phi| : \mathbb{H} \to \mathbb{H}$ is a quantum state it is sufficient to note that $\mathrm{tr}[\mathbf{P}_\phi] = 1$. □

Definition 2.4.3. A quantum state $\rho \in \mathcal{S}(\mathbb{H})$ is a pure state if it is a projection operator onto a one-dimensional subspace of $\mathbb{H}$, i. e., $\rho = \mathbf{P}_\varphi := |\varphi\rangle_{\mathbb{H}}\langle\varphi|$ for some $\varphi \in \mathbb{H}$. All other states are called *mixed states*.

Therefore, a state is a pure state if it can be written as a projection $\mathbf{P}_\varphi : \mathbb{H} \to \mathbb{C}\varphi$ for some $\varphi \in \mathbb{H}$ with $\|\varphi\|_{\mathbb{H}} = 1$, where $\mathbb{C}\varphi := \{c\varphi \mid c \in \mathbb{C}\}$ is the subspace of $\mathbb{H}$ generated by the vector $\varphi \in \mathbb{H}$. In the convex analysis terminology (see Rockfellar [131]), the set of all pure states constitute $\operatorname{extr}(\mathcal{S}(\mathbb{H}))$ (i. e., the set of all extreme points of $\mathcal{S}(\mathbb{H})$).

The following result characterizes the pure states.

Lemma 2.4.4. *A quantum state $\rho \in \mathcal{S}(\mathbb{H})$ is a pure state if and only if it cannot be represented as a nontrivial convex linear combination of elements in $\mathcal{S}(\mathbb{H})$.*

Proof. Suppose that $\mathbf{P}_\psi$ is a convex combination of $\omega_1, \omega_2 \in \mathcal{S}(\mathbb{H})$. This is,

$$\mathbf{P}_\psi = a\omega_1 + (1-a)\omega_2, \quad 0 < a < 1,$$

and let $\mathbb{H} = \mathbb{C}\psi \oplus \mathbb{H}_1$ be the orthogonal sum decomposition of $\mathbb{H}$. Since ω_1 and ω_2 are positive operators, for each $\varphi \in \mathbb{H}_1 = (\mathbb{C}\psi)^\perp$ we have

$$a\langle \omega_1\varphi, \varphi\rangle_{\mathbb{H}} \le \langle \mathbf{P}_\psi\varphi, \varphi\rangle_{\mathbb{H}} = 0,$$

so that $\langle \omega_1\varphi, \varphi\rangle_{\mathbb{H}} = 0$ for all $\varphi \in \mathbb{H}_1$ and by the Cauchy–Schwarz inequality, we get $\omega_1|_{\mathbb{H}_1} = \mathbf{0}$, where $\omega_1|_{\mathbb{H}_1}$ denotes the restriction of ω_1 on the space $\mathbb{H}_1$. Since ω_1 is self-adjoint, it leaves the complementary subspace $\mathbb{C}\psi$ invariant, and from $\operatorname{tr}[\omega_1] = 1$ it follows that $\omega_1 = \mathbf{P}_\psi$. Therefore, $\omega_1 = \omega_2 = \mathbf{P}_\psi$. □

The following result can be found in Chang [23].

Theorem 2.4.5. *A quantum state $\rho \in \mathcal{S}(\mathbb{H})$ has a canonical convex decomposition referred to as a spectral decomposition of ρ of the form*

$$\rho = \sum_{j=1}^{+\infty} \lambda_j \mathbf{P}_j, \tag{2.5}$$

where $\{\lambda_j\}$ is a sequence of nonnegative numbers with $\sum_{j=1}^{+\infty} \lambda_j = 1$ summing to one and $(\mathbf{P}_j)_{j=1}^{+\infty}$ is an orthonormal sequence of one-dimensional projections. If there are infinitely many nonzero terms, then the sum converges with respect to the trace norm $\|\cdot\|_1$.

Proof. This follows from Lemma 2.4.4 and the fact that any convex combination of quantum states is again a quantum state. This proves the theorem. □

In the following example, we construct a quantum state, which is not normal.

Example 2.2. Let $\mathbb{H} = l_2(\mathbb{N}; \mathbb{C})$ be the collection of all square-summable sequences of complex numbers, i. e.,

$$\mathbb{H} = \left\{ \varphi = (x_n)_{n=1}^{+\infty} \mid x_n \in \mathbb{C}, \forall n \in \mathbb{N}, \text{ and } \sum_{n=1}^{+\infty} |x_n|^2 < \infty \right\},$$

and let $\mathcal{A} = l_\infty(\mathbb{N}; \mathbb{C})$ be the collection of all bounded sequences of complex numbers, i. e.,

$$\mathcal{A} = \{\mathbf{a} = (a_n)_{n=1}^{+\infty} \mid a_n \in \mathbb{C}, \forall n \in \mathbb{N}, \ni \|\mathbf{a}\|_\infty \leq K \text{ for some } K \geq 0\}.$$

For $\mathbf{a} = (a_n)_{n=1}^{+\infty} \in \mathcal{A}$ and $\varphi = (x_n)_{n=1}^{+\infty}$ define $\mathbf{a}\varphi = (a_n x_n)_{n=1}^{+\infty}$, i. e., $\mathbf{a}\varphi$ is a pointwise multiplication of $\mathbf{a}$ acting on φ. One can easily show that $\mathcal{A}$ is a commutative von Neumann algebra of bounded linear operators on $\mathbb{H}$. Now introduce a state ρ on $\mathcal{A}$ given by the expression

$$\rho(\mathbf{a}) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{a}_n, \quad \mathbf{a} \in \mathcal{B} \subset \mathcal{A},$$

where

$$\mathcal{B} = \left\{\mathbf{a} \in \mathcal{A} \mid \exists c \in \mathbb{C} \text{ such that } \lim_{n\to\infty} \mathbf{a}_n = c\right\}.$$

It is easy to show that the state ρ constructed above is not normal.

The following lemma repeats some of the result in Proposition 2.4.2. We omit its proof here.

Lemma 2.4.6. *The state space $\mathcal{S}(\mathbb{H})$ is a compact convex subset of $\mathfrak{T}(\mathbb{H})$ under the* weak* *topology.*

2.5 GNS representation

In the following, let $\mathcal{A}$ be a C^*-algebra of bounded linear operators on a nonspecified complex Hilbert space. The purpose of this section is to explore the Gefand–Naiman–Seagal (GNS) representation of $\mathcal{A}$ on some specified complex Hilbert space $\mathbb{H}$ (see, e. g., Chang [23, 24] and Bratteli and Robinson [15] and Takesaki [168].

Definition 2.5.1. A representation of $\mathcal{A}$ is a *-homomorphism $\pi : \mathcal{A} \to \mathfrak{B}(\mathbb{H})$ for some complex Hilbert space $\mathbb{H}$ such that $\pi(\mathbf{ab}) = \pi(\mathbf{a})\pi(\mathbf{b})$ and $\pi(\mathbf{a}^*) = \pi(\mathbf{a})^*$, for all $\mathbf{a}, \mathbf{b} \in \mathcal{A}$. In this case, $\mathbb{H}$ is called the *representation space*. In order to specify the representation space together with a representation, we write $(\pi, \mathbb{H})$ or $\mathbb{H}_\pi$. Two representations $(\pi_1, \mathbb{H}_1)$ and $(\pi_2, \mathbb{H}_2)$ of $\mathcal{A}$ are said to be *unitarily equivalent* if there exists an isometry $\mathbf{U}$ of $\mathbb{H}_1$ onto $\mathbb{H}_2$ such that

$$\mathbf{U}\pi_1(\mathbf{a})\mathbf{U}^* = \pi_2(\mathbf{a}), \quad \mathbf{a} \in \mathcal{A}.$$

In this case, we write $(\pi_1, \mathbb{H}_1) \cong (\pi_2, \mathbb{H}_2)$ or simply $\pi_1 \cong \pi_2$. If $\pi(\mathbf{a}) \neq \mathbf{0}$ for every nonzero $\mathbf{a} \in \mathcal{A}$, then π is called faithful. The set of all representations of a C^*-algebra $\mathcal{A}$ is denoted by $\Pi(\mathcal{A}, \mathfrak{B}(\mathbb{H}))$.

The following result can be found in Takesaki [168] (see also Chang [23, 24]).

Proposition 2.5.2. *Let $(\pi, \mathbb{H})$ be a representation of the C^*-algebra $\mathcal{A}$. Then the following statements are equivalent:*
1. *The closed subspace $[\pi(\mathcal{A})\mathbb{H}]$ spanned by the set $\{\pi(\mathbf{a})\zeta \mid \mathbf{a} \in \mathcal{A}, \zeta \in \mathbb{H}\}$, coincides with the whole space $\mathbb{H}$.*
2. *For any nonzero $\zeta \in \mathbb{H}$, there exists an element $\mathbf{a} \in \mathcal{A}$ with $\pi(\mathbf{a})\zeta \neq 0$.*

Proof. (1) $\Rightarrow$ (2). Suppose that (1) holds. Assume that there exists a nonzero $\zeta \in \mathbb{H}$ such that $\pi(\mathbf{a})\zeta = 0$ for all $\mathbf{a} \in \mathcal{A}$ for contradiction purpose. For any $\eta \in \mathbb{H}$, we have

$$\langle \pi(\mathbf{a})\eta, \zeta \rangle_{\mathbb{H}} = \langle \eta, \pi(\mathbf{a})^* \zeta \rangle_{\mathbb{H}} = \langle \eta, \pi(\mathbf{a}^*)\zeta \rangle_{\mathbb{H}} = 0, \quad \text{since } \mathbf{a} \in \mathcal{A} \text{ implies } \mathbf{a}^* \in \mathcal{A}.$$

Hence, ζ is orthogonal to $[\pi(\mathcal{A})\mathbb{H}]$. By the assumption that $[\pi(\mathcal{A})\mathbb{H}] = \mathbb{H}$, this means that $\zeta = 0$. This is a contradiction. Hence, (2) follows.

(2) $\Rightarrow$ (1). Conversely, suppose that (2) holds. Let ζ be a vector of $\mathbb{H}$ orthogonal to $[\pi(\mathcal{A})\mathbb{H}]$. We then have

$$\begin{aligned} 0 &= \langle \zeta, \pi(\mathbf{a}^*\mathbf{a})\zeta \rangle_{\mathbb{H}} = \langle \zeta, \pi(\mathbf{a}^*)\pi(\mathbf{a})\zeta \rangle_{\mathbb{H}} \\ &= \langle \zeta, \pi^*(\mathbf{a})\pi(\mathbf{a})\zeta \rangle_{\mathbb{H}} = \langle \pi(\mathbf{a})\zeta, \pi(\mathbf{a})\zeta \rangle_{\mathbb{H}}, \quad \mathbf{a} \in \mathcal{A}, \end{aligned}$$

so that $\|\pi(\mathbf{a})\zeta\|_{\mathbb{H}}^2 = 0$ for every $\mathbf{a} \in \mathcal{A}$. By assumption, $\zeta = 0$. Therefore, $([\pi(\mathcal{A})\mathbb{H}])^{\perp} = \{\mathbf{0}\}$. This implies that $[\pi(\mathcal{A})\mathbb{H}] = \mathbb{H}$. Thus, (1) follows. This proves the proposition. □

Definition 2.5.3. Let $(\pi, \mathbb{H})$ be a representation of a C^*-algebra $\mathcal{A}$.
1. The representation $(\pi, \mathbb{H})$ is said to be nondegenerate if for every nonzero vector $v \in \mathbb{H}$, there exists $\mathbf{a} \in \mathcal{A}$ such that $\pi(\mathbf{a})(v) \neq 0$ or equivalently by Proposition 2.5.2 $[\pi(\mathcal{A})\mathbb{H}] = \mathbb{H}$. Otherwise, the closed subspace $[\pi(\mathcal{A})\mathbb{H}]$ is called the *essential space* of π and denoted by $\mathbb{H}(\pi)$.
2. The representation $(\pi, \mathbb{H})$ is called cyclic if there exists a vector $v_\pi \in \mathbb{H}$ such that $\pi(\mathcal{A})v_\pi := \{\pi(\mathbf{a})v_\pi \mid \mathbf{a} \in \mathcal{A}\}$ is dense in $\mathbb{H}$. Such a vector is called a cyclic vector for the representation $(\pi, \mathbb{H})$. In this case, the triple $(\pi, \mathbb{H}, v_\pi)$ will be called a cyclic representation of $\mathcal{A}$.
3. Let $(\pi_i, \mathbb{H}_i)$ for $i \in I$ be a family of representations of C^*-algebra $\mathcal{A}$. Define a representation $\bigoplus_i \pi_i$ on the direct sum $\bigoplus_i \mathbb{H}_i$ by

$$\bigoplus_i \pi_i(\mathbf{a})v = \sum_i \pi_i(\mathbf{a})v_i \quad \text{for } v = \sum_i v_i \text{ and } \mathbf{a} \in \mathcal{A},$$

where the direct sum of Hilbert spaces and operators can be found in Subsection 2.7.2.

2.6 Quantum observables and measurements

The presentation of topics in this section is largely based on results obtained in Chang [22, 23].

2.6.1 Positive operator valued measures

Let $\mathbb{X}$ be a certain locally compact Hausdorff space. Let $\mathbb{H}$ be a complex Hilbert space that represents a certain quantum system, and let $\mathfrak{B}_+(\mathbb{H})$ be the collection of positive bounded linear operators on $\mathbb{H}$, i. e.,

$$\mathfrak{B}_+(\mathbb{H}) = \{\mathbf{a} \in \mathfrak{B}(\mathbb{H}) \mid \mathbf{a} \geq 0\}.$$

Definition 2.6.1. The set function ν on the measurable space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, $\nu : (\mathbb{X}, \mathcal{B}(\mathbb{X})) \to \mathfrak{B}_+(\mathbb{H})$, is said to be a positive operator-valued measure if it satisfies the following conditions:

1. $\nu(E) \leq \nu(F) \leq \nu(\mathbb{X})$ for all $E, F \in \mathcal{B}(\mathbb{X})$ with $E \subseteq F$;
2. $\nu(\bigcup_{n=1}^{+\infty} E_n) = \sum_{n=1}^{+\infty} \nu(E_n)$ for any sequence $(E_n)_{n=1}^{+\infty}$ in $\mathcal{B}(\mathbb{X})$ such that $E_i \cap E_j = \emptyset$ for all $i \neq j$.

The collection of all positive operator-valued measures on the measurable space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ with values in $\mathfrak{B}_+(\mathbb{H})$ will be denoted by $\mathrm{POVM}_{\mathbb{H}}(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ or simply $\mathrm{POVM}_{\mathbb{H}}(\mathbb{X})$.

Definition 2.6.2. If $\nu \in \mathrm{POVM}_{\mathbb{H}}(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ satisfies $\nu(\mathbb{X}) = \mathbf{I}_{\mathbb{H}} \in \mathfrak{B}_+(\mathbb{H})$ ($\mathbf{I}_{\mathbb{H}}$ is the identity operator on $\mathbb{H}$), then ν is said to be a *positive operator-valued probability measure* on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ with values in $\mathfrak{B}_+(\mathbb{H})$.

A positive operator-valued probability measure will sometimes be called a *quantum probability measure*. The class of positive operator-valued probability measures on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ with values in $\mathfrak{B}_+(\mathbb{H})$ is denoted by

$$\mathrm{POVM}_{\mathbb{H}}^1(\mathbb{X}) = \mathrm{POVM}_{\mathbb{H}}^1(\mathbb{X}, \mathcal{B}(\mathbb{X})).$$

If $\nu \in \mathrm{POVM}_{\mathbb{H}}(\mathbb{X})$, then ν induces a finite Borel measure $\mu = \mu_\nu$ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ defined by

$$\mu(E) = \mathrm{tr}[\nu(E)], \quad \forall E \in \mathcal{B}(\mathbb{X}),$$

where $\mathrm{tr}[\cdot] : \mathfrak{B}(\mathbb{H}) \to \mathbb{C}$ is the trace functional. If $\nu \in \mathrm{POVM}_{\mathbb{H}}^1(\mathbb{X})$, then ν induces a probability measure defined on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$.

If $\mathcal{M}(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ denotes the space of finite measures on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, then we can identify $\mathcal{M}(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ with the subset

$$\{\mu \cdot \mathbf{I}_{\mathbb{H}} \mid \mu \in \mathcal{M}(\mathbb{X}, \mathcal{B}(\mathbb{X}))\} \subset \mathrm{POVM}_{\mathbb{H}}(\mathbb{X}).$$

In particular,

$$\{\mu \cdot \mathbf{I}_{\mathbb{H}} \mid \mu \in \mathcal{M}(\mathbb{X}, \mathcal{B}(\mathbb{X})), \mu(\mathbb{X}) = 1\} \subset \mathrm{POVM}^1_{\mathbb{H}}(\mathbb{X}),$$

so that we can consider ordinary probability measures as scalar-valued positive operator-valued probability measures.

In summary, the triple $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \nu)$ is a *quantum probability space* while $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ is a *classical probability space* . Note that ν induces a finite Borel measure $\mu = \mu_\nu$ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$.

2.6.2 Quantum observables

To formulate von Neumann's postulate 3, we now give a formal definition of a quantum observable below (see Chang [23]).

Definition 2.6.3. The triple $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mathbf{a})$ is said to be a *quantum observable* if (i) $\mathbb{X}$ is a locally compact Hausdorff space; (ii) $\mathcal{B}(\mathbb{X})$ is the Borel σ-algebra of subsets of $\mathbb{X}$ and (iii) $\mathbf{a}$ is a positive self-adjoint operator valued measure $\mathbf{a} : (\mathbb{X}, \mathcal{B}(\mathbb{X})) \to \mathfrak{B}_+(\mathbb{H})$ such that $\mathbf{a}(E)$ is a positive self-adjoint operator on the complex Hilbert space $\mathbb{H}$ for every $E \in \mathcal{B}(\mathbb{X})$ that satisfies:
1. $\mathbf{0} \le \mathbf{a}(E) \le \mathbf{a}(\mathbb{X})$;
2. $\mathbf{a}(\mathbb{X}) = \tau$, where $\tau : \mathbb{H} \to \mathbb{C}$ is a bounded linear functional on $\mathbb{H}$ such that $\tau(\phi) = \|\phi\|_{\mathbb{H}}$;
3. $\mathbf{a}(\bigcup_{n=1}^{+\infty} E_n) = \sum_{n=1}^{+\infty} \mathbf{a}(E_n)$ for any sequence $\{E_n, n = 1, 2, \ldots\}$ of pairwise disjoint sets in $\mathcal{B}(\mathbb{X})$, where the summation in right-hand side is the σ-weakly convergence.

In this case, the measurable space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ is said to be the value space of **a**.

The collection of bounded quantum observables will be denoted by $\mathcal{O}_{\mathbb{H}}(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. As described in Chang [22], a "quantum observable" is the quantum physicist's word for real random variable that describes a physical quantity (such as position, velocity, momentum, angular momentum, energy, etc.) of a quantum system that plays a central role in quantum mechanics. They are mathematical representations of physical quantities that can (in principle) be measured. However, arbitrary nonreal elements (or nonself-adjoint operators) do not represent in general complex random variables. Nonreal (i. e., complex) quantum random variables correspond to *normal* elements $\mathbf{a} \in \mathcal{A}$, which commutes with their adjoint, i. e., $\mathbf{a}(E)\mathbf{a}^*(E) = \mathbf{a}^*(E)\mathbf{a}(E)$ for all $E \in \mathcal{B}(\mathbb{X})$. To avoid unnecessarily confusion, we often take $\mathbb{X}$ to be $\mathbb{R}$ for simplicity. In this case, all quantum observables are assumed to be real-valued.

The definition of quantum probability space is given below.

Definition 2.6.4 (Quantum probability space). A quantum probability space is a pair $(\mathcal{A}, \rho)$, where $\mathcal{A}$ is a von Neumann algebra of bounded linear operators on a complex Hilbert space $\mathbb{H}$ and ρ is a normal (i. e., σ-weakly continuous) state. The events in $(\mathcal{A}, \rho)$ are the orthogonal projections $\mathbf{p} \in \mathcal{A}$. The probability that $\mathbf{p}$ occurs is $\rho(\mathbf{p})$.

2.6.3 Quantum measurements

Formulation of von Neumann's postulate 4 is as follows. Recall that a quantum state ρ is a positive trace-class operator such that $\operatorname{tr}(\rho) = 1$ and a quantum observable $\mathbf{a}$ is a self-adjoint operator-valued map defined on the real measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. By the von Neumann spectral theorem (Theorem 1.7.4), there exists a projection-valued measure $\mu_{\mathbf{a}}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathbf{a}(E) = \int_E \lambda \mu_{\mathbf{a}}(d\lambda)$ for all $E \in \mathcal{B}(\mathbb{R})$. The probability $P(\rho, \mathbf{a}, E)$ that in the quantum state ρ the quantum observable $\mathbf{a}$ should take values in $E \in \mathcal{B}(\mathbb{R})$ is given by $P(\rho, \mathbf{a}, E) = \operatorname{tr}[\rho\mu_{\mathbf{a}}(E)]$.

A *measurement* of the real quantum observable $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{a})$ (or simply $\mathbf{a}$) is a physical procedure or experiment that produces numerical results related to $\mathbf{a}$. A process of measurement is the map $(\mathbf{a}, \rho) \mapsto \mu_{\mathbf{a}}$ from $\mathcal{A} \times \mathcal{S}(\mathcal{A})$ to $\mathcal{P}(\mathbb{R})$ (where $\mathcal{P}(\mathbb{R})$ is the space of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which to every observable $\mathbf{a} \in \mathcal{A}$ and state $\rho \in \mathcal{S}(\mathcal{A})$ assigns a probability measure μ on the Borel measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For every Borel subset $E \in \mathcal{B}(\mathbb{R})$, the quantity $0 \leq \mu_{\mathbf{a}}(E) \leq 1$ is the probability that for a quantum system in the state ρ the result of a measurement of the observable $\mathbf{a}$ belongs to E. The expectation value (the mean-value) of the observable $\mathbf{a} \in \mathcal{A}$ is $\int_{\infty}^{\infty} \lambda d\mu_{\mathbf{a}}(\lambda)$, where $\mu_{\mathbf{a}}(\lambda) = \mu_{\mathbf{a}}(]-\infty, \lambda[)$ is a distribution function for the probability measure $\mu_{\mathbf{a}}$.

In any given measurement of the observable $\mathbf{a}$, the allowable results a take values in $\sigma(\mathbf{a})$, the spectrum of $\mathbf{a}$. Given the state ρ, the value $a \in \sigma(\mathbf{a})$ is observed with probability $\operatorname{tr}(\rho\mathbf{P}_{\psi(a)})$, where $\mathbf{P}_{\psi(a)}$ or simply $\mathbf{P}_a$ is the one-dimensional vector space generated by the eigenvector $\psi(a)$ associated to the eigenvalue a of $\mathbf{a}$. Consequently, the expectation of the observable $\mathbf{a}$ is given by $\mathbf{E}_\rho(\mathbf{a}) = \operatorname{tr}[\rho\mathbf{a}]$.

Suppose that a measurement of the observable $\mathbf{a}$ gives rise to the observation $a \in \sigma(\mathbf{a})$. Then we must condition that state in order to predict the outcomes of subsequent measurements, by updating the state ρ using

$$\rho \mapsto \rho'[a] = \frac{\mathbf{P}_a \rho \mathbf{P}_a}{\operatorname{tr}(\rho\mathbf{P}_a)}. \tag{2.6}$$

This is the so-called *back-action* of a quantum measurement.

Let $\mathcal{B}(\mathbb{R})$ be the σ-algebra of Borel subsets of $\mathbb{R}$ (see Wheeden and Zygmund [177] for a definition of Borel σ-algebra). Recall that $\mathfrak{L}_p(\mathbb{H})$ is the collection of projection operators on $\mathbb{H}$ (see Section 1.6 for the definition and properties of projection operators). In the following, we explore the concept of *projection-valued measures*.

Definition 2.6.5. A projection-valued measure on the Borel measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a mapping $\mathbf{P} : \mathcal{B}(\mathbb{R}) \to \mathfrak{L}_p(\mathbb{H})$ satisfying the following properties:

1. For every $E \in \mathcal{B}(\mathbb{R})$, $\mathbf{P}(E)$ is an orthogonal projection, i. e., $\mathbf{P}(E) = \mathbf{P}^2(E) := (\mathbf{P}(E))^2$ and $\mathbf{P}(E)$ is self-adjoint, i,e., $\mathbf{P}(E) = \mathbf{P}^*(E) := (\mathbf{P}(E))^*$.
2. $\mathbf{P}(\emptyset) = \mathbf{0}_{\mathbb{H}}$ (the zero operator on $\mathbb{H}$) and $\mathbf{P}(\mathbb{R}) = \mathbf{I}_{\mathbb{H}}$ (the identity operator on $\mathbb{H}$).
3. For every disjoint sequence $(E_n)_{n=1}^{\infty} \subset \mathcal{B}(\mathbb{R})$ such that $E_n \cap E_m = \emptyset$ for all $n \neq m$, we have

$$\lim_{n\to\infty} \left\| \mathbf{P}(\cup_{n=1}^{\infty} E_n)\phi - \sum_{i=1}^{n} \mathbf{P}(E_i)\phi \right\|_{\mathbb{H}}, \quad \forall \phi \in \mathbb{H}.$$

Remark 2.4. We make the following observations for extension of projection-valued measures to $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$, $N \in \mathbb{N}$, as follows:

1. A projection-valued measure on $\mathcal{B}(\mathbb{R}^N)$ is a mapping $\mathbf{P} : \mathcal{B}(\mathbb{R}^N) \to \mathfrak{L}_p(\mathbb{H})$ satisfying conditions similar to properties (1)–(3) in Definition 2.6.5.
2. If follows from Definition 2.6.5 that

$$\mathbf{P}(E_1)\mathbf{P}(E_2) = \mathbf{P}(E_1 \cap E_2), \quad \forall E_1, E_2 \in \mathcal{B}(\mathbb{R}^N).$$

Definition 2.6.6. Let $\mathbf{P} : \mathcal{B}(\mathbb{R}) \to \mathfrak{L}_p(\mathbb{H})$ be a projection-valued measure. We associate $\mathbf{P}$ with a projection-valued function $\mathbf{P} : \mathbb{R} \to \mathfrak{L}_p(\mathbb{H})$ defined by

$$\mathbf{P}(\lambda) = \mathbf{P}(]-\infty, \lambda]), \quad \forall \lambda \in \mathbb{R}.$$

The projection-valued function defined above will be called the *projection-valued resolution of identity*.

Remark 2.5. A projection-valued resolution of identity $\mathbf{P} : \mathbb{R} \to \mathfrak{L}_p(\mathbb{H})$ can be characterized by the following properties:

1. $\mathbf{P}(\lambda)\mathbf{P}(\mu) = \mathbf{P}(\min\{\lambda, \mu\})$.
2. $\lim_{\lambda\to-\infty} \mathbf{P}(\lambda) = \mathbf{0}$ and $\lim_{\lambda\to+\infty} \mathbf{P}(\lambda) = \mathbf{I}$.
3. $\lim_{\mu\uparrow\lambda} \mathbf{P}(\mu) = \mathbf{P}(\lambda)$

For every $\varphi \in \mathbb{H}$, the projection-valued measure $\mathbf{P} : \mathcal{B}(\mathbb{R}) \to \mathfrak{L}_p(\mathbb{H})$ (resp., the resolution of identity $\mathbf{P} : \mathbb{R} \to \mathfrak{L}_p(\mathbb{H})$) define a real-valued measure $\langle \mathbf{P}(\cdot)\varphi, \varphi\rangle_{\mathbb{H}}$ on $\mathcal{B}(\mathbb{R})$ (resp., a nondecreasing function $\langle \mathbf{P}(\cdot)\varphi, \varphi\rangle_{\mathbb{H}}$ defined on $\mathbb{R}$). In the case where $\|\varphi\|_{\mathbb{H}} = 1$, the real-valued measure $\langle \mathbf{P}(\cdot)\varphi, \varphi\rangle_{\mathbb{H}}$ becomes a probability measure defined on $\mathcal{B}(\mathbb{R})$ and the nondecreasing function $\langle \mathbf{P}(\cdot)\varphi, \varphi\rangle_{\mathbb{H}}$ becomes a (probability) distribution function defined on $\mathbb{R}$. The real-valued measure $\langle \mathbf{P}(\cdot)\varphi, \varphi\rangle_{\mathbb{H}}$ (resp., the nondecreasing function $\langle \mathbf{P}(\cdot)\varphi, \varphi\rangle_{\mathbb{H}}$) can be extended to a complex-valued measure (a complex linear combination of measures) defined on $\mathcal{B}(\mathbb{R})$ (resp., complex-valued function on $\mathbb{R}$) by the following polarization identity:

$$\langle \mathbf{P}(\cdot)\varphi, \phi\rangle_{\mathbb{H}} = \frac{1}{4}\{\langle \mathbf{P}(\cdot)(\varphi+\phi), \varphi+\phi\rangle_{\mathbb{H}} - \langle \mathbf{P}(\cdot)\varphi - \phi, \varphi - \phi\rangle_{\mathbb{H}} \\ + i\langle \mathbf{P}(\cdot)(\varphi + i\phi), \varphi + i\phi\rangle_{\mathbb{H}} - i\langle \mathbf{P}(\cdot)(\varphi - i\phi), \varphi - i\phi\rangle_{\mathbb{H}})\}.$$

A measurable function f defined on the Borel measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is said to be finite almost everywhere (a. e.) with respect to the projection-valued measure $\mathbf{P}$ if it is finite a. e. with respect to all measures $\langle \mathbf{P}(\cdot)\varphi, \varphi\rangle_{\mathbb{H}}$, $\varphi \in \mathbb{H}$. If the Hilbert space $\mathbb{H}$ is separable, a theorem of von Neumann states that for every projection-valued measure $\mathbf{P} : \mathcal{B}(\mathbb{R}) \to \mathfrak{L}_p(\mathbb{H})$ there exists $\varphi \in \mathbb{H}$ such that a function f is finite a. e. with respect to $\mathbf{P}$ if and only if it is finite a. e. with respect to the measure $\langle \mathbf{P}(\cdot)\varphi, \varphi\rangle_{\mathbb{H}}$.

2.7 Tensor products and direct sums

In the following subsections, we review the concepts of tensor products and direct sums of Hilbert spaces and operators, which will be used in the remainder of this chapter and beyond.

2.7.1 Tensor products of Hilbert spaces and operators

In the following, we illustrate the concept and construction of tensor product $\mathbb{H}_A \otimes \mathbb{H}_B$ for only two Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$. For notational simply, we often write $\mathbb{H}_A \otimes \mathbb{H}_B$ as $\mathbb{H}_{AB}$, $\langle\cdot,\cdot\rangle_{\mathbb{H}_A}$ as $\langle\cdot,\cdot\rangle_A$, $\langle\cdot,\cdot\rangle_{\mathbb{H}_{AB}}$ as $\langle\cdot,\cdot\rangle_{AB}$, etc. The concepts and constructions can be easily extended to more than two Hilbert spaces.

Recall that the algebraic *tensor product* of $\mathbb{H}_A$ with $\mathbb{H}_B$ and to be denoted as $\mathbb{H}_A \odot \mathbb{H}_B$ consists of elements $\phi \otimes \varphi$ ($\phi \in \mathbb{H}_A$ and $\varphi \in \mathbb{H}_B$) and is equipped with the inner product

$$\langle\cdot,\cdot\rangle_{\mathbb{H}_A \odot \mathbb{H}_B} : (\mathbb{H}_A \odot \mathbb{H}_B) \times (\mathbb{H}_{\mathbb{A}} \odot \mathbb{H}_B) \to \mathbb{C}$$

defined by

$$\langle \phi_1 \otimes \varphi_1, \phi_2 \otimes \varphi_2\rangle_{\mathbb{H}_A \odot \mathbb{H}_B} = \langle \phi_1, \phi_2\rangle_A \langle \varphi_1, \varphi_2\rangle_B, \phi_i \in \mathbb{H}_A \text{ and } \varphi_i \in \mathbb{H}_B \text{ for } i = 1, 2.$$

Note that $\mathbb{H}_A \odot \mathbb{H}_B$ is not yet a Hilbert space with respect to the inner product defined above. The completion of $\mathbb{H}_A \odot \mathbb{H}_B$ under the norm

$$\|\cdot\|_{\mathbb{H}_A \odot \mathbb{H}_B} := \sqrt{\langle\cdot,\cdot\rangle_{\mathbb{H}_A \odot \mathbb{H}_B}}$$

will be denoted by $\mathbb{H}_A \otimes \mathbb{H}_B$ (or $\mathbb{H}_{AB}$ for notational simplicity). It can be easily proved that the composite space $\mathbb{H}_{AB}$ is a complex Hilbert space equipped with the inner product

$$\langle\cdot,\cdot\rangle_{AB} = \langle\cdot,\cdot\rangle_{\mathbb{H}_A\odot\mathbb{H}_B}$$

and the Hilbertian norm $\|\cdot\|_{AB}$, etc. The elements $\phi\otimes\varphi$ of the tensor product $\mathbb{H}_A\otimes\mathbb{H}_B$ are linear with respect to both arguments and satisfy the following distributive law:

$$\begin{aligned} c(\phi\otimes\varphi) &= (c\phi)\otimes\varphi = \phi\otimes(c\varphi), \\ (\phi_1+\phi_2)\otimes\varphi &= \phi_1\otimes\varphi+\phi_2\otimes\varphi, \\ \phi\otimes(\varphi_1+\varphi_2) &= \phi\otimes\varphi_1+\phi\otimes\varphi_2, \end{aligned}$$

for every $\phi,\phi_1,\phi_2\in\mathbb{H}_A$, $\varphi,\varphi_1,\varphi_2\in\mathbb{H}_B$ and $c\in\mathbb{C}$.

It is clear that if $\{\phi_i\}_{i=1}^{+\infty}$ and $\{\varphi_i\}_{i=1}^{+\infty}$ are orthonormal bases of $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively, then $\{\phi_i\otimes\varphi_j\}_{i,j=1}^{+\infty}$ is an orthonormal basis for $\mathbb{H}_{AB} := \mathbb{H}_A\otimes\mathbb{H}_B$. Every $\zeta\in\mathbb{H}_{AB}$ can be expanded in terms of $\{\phi_i\otimes\varphi_j\}_{i,j=1}^{+\infty}$ as $\zeta=\sum_{i,j}\zeta_{ij}(\phi_i\otimes\varphi_j)$ with $\zeta_{ij}=\langle\phi_i\otimes\varphi_j,\zeta\rangle_{AB}$. This procedure works for an arbitrary number of tensor factors. However, if we have exactly a twofold tensor product, there is a more economic way to expand ζ, called the *Schmidt decomposition* in which only diagonal terms of the form $\phi_i\otimes\varphi_i$ appear. The Schmidt decomposition will be explored in detail in Section 10.2.2.

Definition 2.7.1 (Tensor products of linear operators). The tensor product $\mathbf{A}\otimes\mathbf{B}$ of two linear operators $\mathbf{A}\in\mathfrak{L}(\mathbb{H})$ and $\mathbf{B}\in\mathfrak{L}(\mathbb{K})$ is the unique linear operator on $\mathbb{H}\otimes\mathbb{K}$ defined by

$$(\mathbf{A}\otimes\mathbf{B})(\phi\otimes\psi)=\mathbf{A}(\phi)\otimes\mathbf{B}(\psi),\quad \forall\phi\in\mathbb{H}\text{ and }\forall\psi\in\mathbb{K}.$$

The tensor product $\mathbf{A}_1\otimes\mathbf{A}_2\otimes\cdots\otimes\mathbf{A}_n$ of $\mathbf{A}_i\in\mathfrak{L}(\mathbb{H}_i)$ for $i=1,2,\ldots,n$ can be similarly defined.

Definition 2.7.2 (Tensor products of vector spaces of linear operators). Let $\mathcal{A}$ and $\mathcal{B}$ be two vector spaces of operators over $\mathbb{C}$. The tensor product of $\mathcal{A}$ and $\mathcal{B}$, denoted by $\mathcal{A}\otimes\mathcal{B}$, is the vector space spanned by the elementary tensors $\mathbf{a}\otimes\mathbf{b}$, $\mathbf{a}\in\mathcal{A}$ and $\mathbf{b}\in\mathcal{B}$, satisfying the following relations:

$$\begin{aligned} (\mathbf{a}+\mathbf{a}')\otimes\mathbf{b} &= \mathbf{a}\otimes\mathbf{b}+\mathbf{a}'\otimes\mathbf{b}; \\ \mathbf{a}\otimes(\mathbf{b}+\mathbf{b}') &= \mathbf{a}\otimes\mathbf{b}+\mathbf{a}\otimes\mathbf{b}'; \\ (\lambda\mathbf{a})\otimes\mathbf{b} &= \mathbf{a}\otimes(\lambda\mathbf{b})=\lambda(\mathbf{a}\otimes\mathbf{b}), \end{aligned}\tag{2.7}$$

for all $\mathbf{a},\mathbf{a}'\in\mathcal{A}$, $\mathbf{b},\mathbf{b}'\in\mathcal{B}$ and $\lambda\in\mathbb{C}$.

Remark 2.6. We have the following results regarding tensor products:
1. (Tensor products of complex algebras). Let $\mathcal{A}$ and $\mathcal{B}$ be two complex algebras of linear operators. Then the vector space $\mathcal{A}\otimes\mathcal{B}$ becomes a complex algebra if we define

$$(\mathbf{a} \otimes \mathbf{b})(\mathbf{a}' \otimes \mathbf{b}') = \mathbf{aa}' \otimes \mathbf{bb}'$$

and extend linearly to $\mathcal{A} \otimes \mathcal{B}$.

2. (Tensor products of *-algebras). Let $\mathcal{A}$ and $\mathcal{B}$ be two *-algebras. Then the complex algebra $\mathcal{A} \otimes \mathcal{B}$ becomes a *-algebra if we define

$$\sum_i (\mathbf{a}_i \otimes \mathbf{b}_i)^* = \sum_i \mathbf{a}_i^* \otimes \mathbf{b}_i^*.$$

3. (Tensor products of C^*-algebras). Let $\mathcal{A} \subseteq \mathfrak{B}(\mathbb{H}_A)$ and $\mathcal{B} \subseteq \mathfrak{B}(\mathbb{H}_B)$ be C^*-algebras of bounded linear operators. Then an element $\sum_i \mathbf{a}_i \otimes \mathbf{b}_i$ of the *-algebra $\mathcal{A} \otimes \mathcal{B}$ can be viewed as an operator on the inner product space $\mathbb{H}_{AB}$ if we set

$$\Big(\sum_i \mathbf{a}_i \otimes \mathbf{b}_i\Big)\Big(\sum_j x_j \otimes y_j\Big) = \sum_{i,j} \mathbf{a}_i x_j \otimes \mathbf{b}_i y_j.$$

With respect to the operator norm $\|\cdot\|_\infty$ on $\mathfrak{B}(\mathbb{H}_{AB})$, $\mathcal{A} \otimes \mathcal{B}$ becomes a *-algebra with the operator norm (also denoted by $\|\cdot\|_\infty$) satisfying

$$\|\mathbf{uv}\|_\infty \le \|\mathbf{u}\|_\infty \|[\mathbf{v}\|_\infty \quad \text{and} \quad \|\mathbf{u}^*\mathbf{u}\|_\infty = \|\mathbf{u}\|_\infty^2.$$

Hence, the completion of $\mathcal{A} \otimes \mathcal{B}$ becomes a C^*-algebra.

The following provides an example of the tensor product of two bounded linear operators.

Example 2.3. Consider the $n \times n$ matrix A (considered as a linear operator on $\mathbb{H}_A = \mathbb{C}^n$ the $m \times m$ matrix B (considered as a linear operator on $\mathbb{H}_B = \mathbb{C}^m$, which are explicitly written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mm} \end{bmatrix}.$$

Then $A \otimes B$ is an $nm \times nm$ matrix (considered as a linear operator on $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B = \mathbb{C}^{nm}$) and can be written as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1B} & a_{n2}B & \cdots & a_{nn}B \end{bmatrix},$$

where $a_{ij}B$ (an $m \times m$ matrix) is the scalar product of a_{ij} and B.

2.7.2 Direct sum of Hilbert spaces and operators

In this section, we briefly review direct sum of Hilbert spaces and operators, which will be involved in proofs of many results in this and the next chapter.

Consider a family of complex Hilbert spaces $\{\mathbb{H}_i\}_{i\in\mathbb{I}}$, where $\mathbb{I}$ is a finite or countably infinite index set. We define the direct sum of this family of Hilbert spaces as

$$\bigoplus_{i\in\mathbb{I}} \mathbb{H}_i = \left\{(\psi_i)_{i\in\mathbb{I}} \in \prod_{i\in\mathbb{I}} \mathbb{H}_i \mid \psi_i \in \mathbb{H}_i, i \in \mathbb{I}, \text{ and } \sum_{i\in\mathbb{I}} \|\psi_i\|^2_{\mathbb{H}_i} < +\infty\right\}, \tag{2.8}$$

where $\prod_{i\in\mathbb{I}}$ denotes the Cartesian product and $\sum_{i\in\mathbb{I}}$ denotes the generalized sum. The direct sum $\bigoplus_{i\in\mathbb{I}} \mathbb{H}_{i\in\mathbb{I}}$ is a complex Hilbert space equipped with the inner produce $\langle\cdot,\cdot\rangle_{\bigoplus_{i\in\mathbb{I}} \mathbb{H}_i}$ defined by

$$\langle(\phi_i)_{i\in\mathbb{I}}, (\psi_i)_{i\in\mathbb{I}}\rangle_{\bigoplus_{i\in\mathbb{I}} \mathbb{H}_i} = \sum_{i\in\mathbb{I}} \langle\phi_i, \psi_i\rangle_{\mathbb{H}_i}, \quad \forall (\phi_i)_{i\in\mathbb{I}}, (\psi_i)_{i\in\mathbb{I}} \in \bigoplus_{i\in\mathbb{I}} \mathbb{H}_i. \tag{2.9}$$

If $\mathbb{H}_i = \mathbb{H}$ for all $i \in \mathbb{I}$, one can easily prove that for each $k \in \mathbb{N}$, the k-fold direct sum

$$\mathbb{H}^{\bigoplus k} = \underbrace{\mathbb{H} \oplus \mathbb{H} \oplus \cdots \oplus \mathbb{H}}_{k \text{ folds}}$$

is isomorphic to $\mathbb{C}^k \otimes \mathbb{H}$. The isomorphism between $\mathbb{H}^{\oplus k}$ and $\mathbb{C}^k \otimes \mathbb{H}$ is relevant in the context of superpositions of quantum states and entanglement.

Let $(\mathbb{H}_n)_{n=1}^{+\infty}$ be a sequence of complex Hilbert spaces. Let $\bigoplus_{n=1}^{+\infty} \mathbb{H}_n$ be the direct sum of $(\mathbb{H}_n)_{n=1}^{+\infty}$ (see (2.8) and (2.9) for the definition of direct sum of Hilbert spaces and related properties).

For each $n \in \mathbb{N}$, let $\mathbf{T}_n \in \mathfrak{B}(\mathbb{H}_n)$ be a bounded linear operator on $\mathbb{H}_n$. We define the direct sum $\mathbf{T} = \bigoplus_{n=1}^{+\infty}\mathbf{T}_n : \bigoplus_{n=1}^{+\infty} \mathbb{H}_n \to \bigoplus_{n=1}^{+\infty} \mathbb{H}_n$ by

$$\mathbf{T}(\varphi) = \big(\mathbf{T}_n(\varphi_n)\big)_{n=1}^{+\infty}, \quad \forall \varphi = (\varphi_n)_{n=1}^{+\infty}, \tag{2.10}$$

where $\varphi_n \in \mathbb{H}_n$.

We have the following result regarding $\mathbf{T} : \bigoplus_{n=1}^{+\infty} \mathbb{H}_n \to \bigoplus_{n=1}^{+\infty} \mathbb{H}_n$.

Theorem 2.7.3. *Assume that $\mathbf{T}_n \in \mathfrak{B}(\mathbb{H}_n)$ for each $n \in \mathbb{N}$ such that*

$$\sup_{n\in\mathbb{N}} \|\mathbf{T}_n\|_\infty < +\infty,$$

where $\|\mathbf{T}_n\|_\infty$ is the operator norm of $\mathbf{T}_n$. Then the operator $\mathbf{T}$ defined in (2.10) is a bounded linear operator on $\bigoplus_{n=1}^{+\infty} \mathbb{H}_n$ with the operator norm also denoted by $\|\cdot\|_\infty$ (for notational simplicity) and defined by

$$\|\mathbf{T}\|_\infty = \sup_{n\in\mathbb{N}} \|\mathbf{T}_n\|_\infty.$$

Proof. It is clear that the operator $\mathbf{T} : \bigoplus_{n=1}^{+\infty} \mathbb{H}_n \to \bigoplus_{n=1}^{+\infty} \mathbb{H}_n$ is a linear operator. To show that $\mathbf{T}$ is bounded on $\bigoplus_{n=1}^{+\infty} \mathbb{H}_n$, we let $c = \sup_{n\in\mathbb{N}} \|\mathbf{T}_n\|_\infty$. By assumption, $c < +\infty$. Let $\varphi = (\varphi_n)_{n=1}^{+\infty}$ be an arbitrary vector in $\bigoplus_{n=1}^{+\infty} \mathbb{H}_n$. Then

$$\begin{aligned}
\|\mathbf{T}\varphi\|^2_{\bigoplus_{n=1}^{+\infty}\mathbb{H}_n} &= \|(\mathbf{T}_n\varphi_n)_{n=1}^{+\infty}\|^2_{\bigoplus_{n=1}^{+\infty}\mathbb{H}_n} \quad \text{(by the definition of } \mathbf{T})\\
&= \sum_{n=1}^{+\infty} \|\mathbf{T}_n\varphi_n\|^2_{\mathbb{H}_n} \quad \text{(by the definition of } \|\cdot\|_{\mathbb{H}})\\
&\le \sum_{n=1}^{+\infty} \|\mathbf{T}_n\|^2_\infty \|\varphi_n\|^2_{\mathbb{H}_n}\\
&\le \sum_{n=1}^{+\infty} c^2\|\varphi_n\|^2_{\mathbb{H}_n} = c^2 \sum_{n=1}^{+\infty} \|\varphi_n\|^2_{\mathbb{H}_n} = c^2\|\varphi\|_{\bigoplus_{n=1}^{+\infty}\mathbb{H}_n}.
\end{aligned}$$

In summary, we have $\|\mathbf{T}\varphi\|_{\bigoplus_{n=1}^{+\infty}\mathbb{H}_n}| \le c\|\varphi\|_{\bigoplus_{n=1}^{+\infty}\mathbb{H}_n}$ for all $\varphi = (\varphi_n)_{n=1}^{+\infty} \in \bigoplus_{n=1}^{+\infty}\mathbb{H}_n$. Therefore, the operator $\mathbf{T}$ is a bounded linear operator on $\bigoplus_{n=1}^{+\infty}\mathbb{H}_n$. From the above derivation, it is clear that the operator norm $\|\mathbf{T}\|_\infty$ defined by

$$\|\mathbf{T}\|_\infty = \sup_{\|\varphi\|_{\bigoplus_{n=1}^{+\infty}\mathbb{H}_n}=1} \|\mathbf{T}\varphi\|_{\bigoplus_{n=1}^{+\infty}\mathbb{H}_n} \le c\|\mathbf{T}\varphi\|_{\bigoplus_{n=1}^{+\infty}\mathbb{H}_n}.$$

On the other hand, $\|\mathbf{T}\|_\infty = c := \sup_{n\in\mathbb{N}} \|\mathbf{T}_n\|_\infty$ by choosing φ_1 as a unit vector in $\mathbb{H}_1$ and $\varphi_n = 0$ for $n = 2, 3, \cdots$. Therefore, $\|\mathbf{T}\|_\infty = \sup_{n\in\mathbb{N}} \|\mathbf{T}_n\|_\infty$. This proves the theorem. □

The following example gives an explicit computation of the direct sum of the two matrices A and B.

Example 2.4. Consider the $n\times n$ matrix A (considered as a linear operator on $\mathbb{H}_A = \mathbb{C}^n$) the $m \times m$ matrix B (considered as a linear operator on $\mathbb{H}_B = \mathbb{C}^m$), which are explicitly written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m}\\ b_{21} & b_{22} & \cdots & b_{2m}\\ \vdots & \vdots & \vdots & \vdots\\ b_{m1} & b_{m2} & \cdots & b_{mm} \end{bmatrix}.$$

Then $A \oplus B$ is an $(n + m) \times (n + m)$ matrix (considered as a linear operator on $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B = \mathbb{C}^{nm}$) and can be written as the block matrix

$$A \oplus B = \begin{bmatrix} A & 0\\ 0 & B \end{bmatrix},$$

where 0 above mean zero matrices of appropriate sizes.

2.8 Composite quantum systems

This section is devoted to formulation of von Neumann's postulate 1, where a composite quantum system is to be expressed as tensor products of its component systems.

2.8.1 Composite system as tensor products

The Hilbert space $\mathbb{H}$ representing a composite quantum system that consists of n subsystems can be written as tensor product of n Hilbert spaces of its component systems described by $\mathbb{H}_1, \ldots, \mathbb{H}_n$, i. e., $\mathbb{H} = \mathbb{H}_1 \otimes \ldots \otimes \mathbb{H}_n$. In particular, for a bipartite quantum system $\mathbb{H}$ that consists of component systems $\mathbb{H}_A$ and $\mathbb{H}_B$, we write $\mathbb{H} = \mathbb{H}_A \otimes \mathbb{H}_B$ or simply $\mathbb{H}_{AB}$ where the rules/properties of tensor product of Hilbert spaces and operators reviewed in Subsection 2.7.1 are applicable.

2.8.2 Partial traces

In this subsection, we are interested in computing partial traces of a bounded linear operator defined on a composite Hilbert space over each of its component Hilbert spaces.

Recall from (1.23) that $\mathrm{tr}[\mathbf{T}]$, the trace of $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$, is defined as $\mathrm{tr}[\mathbf{T}] = \sum_{i=1}^{+\infty} \langle \phi_i, \mathbf{T}\phi_i \rangle_{\mathbb{H}}$, where $\{\phi_i\}_{i=1}^{+\infty}$ is any orthonormal basis of $\mathbb{H}$.

Assume that a quantum system A interacts with another quantum system B. In this case, the composite Hilbert space $\mathbb{H}$ can be decomposed in the form $\mathbb{H} = \mathbb{H}_A \otimes \mathbb{H}_B := \mathbb{H}_{AB}$, where subsystems A and B are represented by the complex Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. The partial trace operators $\mathrm{tr}_{\mathbb{H}_A}[\cdots] : \mathfrak{T}(\mathbb{H}_{AB}) \to \mathfrak{T}(\mathbb{H}_B)$ and $\mathrm{tr}_{\mathbb{H}_B}[\cdots] : \mathfrak{T}(\mathbb{H}_{AB}) \to \mathfrak{T}(\mathbb{H}_A)$ can be defined abstractly as

$$\mathrm{tr}_{\mathbb{H}_B}[\mathbf{T}_A \otimes \mathbf{T}_B] = \mathbf{T}_A\, \mathrm{tr}[\mathbf{T}_B] \quad \text{and} \quad \mathrm{tr}_{\mathbb{H}_A}[\mathbf{T}_A \otimes \mathbf{T}_B] = \mathrm{tr}[\mathbf{T}_A]\mathbf{T}_B,$$

for all $\mathbf{T}_A \in \mathfrak{T}(\mathbb{H}_A)$ and for all $\mathbf{T}_B \in \mathfrak{T}(\mathbb{H}_B)$.

For notational simplicity, we often write $\mathrm{tr}_{\mathbb{H}_A}[\cdots]$ as $\mathrm{tr}_A[\cdot]$, $\mathrm{tr}_{\mathbb{H}_B}[\cdots]$ as $\mathrm{tr}_B[\cdots]$, etc.

For any trace-class operator $\mathbf{T} \in \mathfrak{T}(\mathbb{H}_{AB})$, we are interested in computing its partial trace $\mathrm{tr}_A[\mathbf{T}](\in \mathfrak{T}(\mathbb{H}_B))$ taken over the Hilbert space $\mathbb{H}_A$ and partial trace $\mathrm{tr}_B[\mathbf{T}](\in \mathfrak{T}(\mathbb{H}_A))$ taken over $\mathbb{H}_B$. The concept of partial trace of trace-class operators on the system that consists of any finite number of subsystems can be similarly extended.

We illustrate how to compute $\mathrm{tr}_B[\mathbf{T}]$ below (note that the partial trace $\mathrm{tr}_A[\mathbf{T}]$ can be similarly computed). Let $(\psi_i)_{i=1}^{+\infty}$ be an orthonormal basis for $\mathbb{H}_B$. We consider the following isometric isomorphism:

$$\bigoplus_{i=1}^{+\infty} (\mathbb{H}_A \otimes \mathbb{C}\psi_i) \mapsto \mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B,$$

where $\mathbb{H}_A \otimes \mathbb{C}\psi_i$ is defined as

$$\mathbb{H}_A \otimes \mathbb{C}\psi_i = \{\phi \otimes c\psi_i | \phi \in \mathbb{H}_A, c \in \mathbb{C}\}$$

and $\bigoplus_{n=1}^{+\infty}(\mathbb{H}_A \otimes \mathbb{C}\psi_i)$ is the direct sum of $(\mathbb{H}_A \otimes \mathbb{C}\psi_i)_{n=1}^{+\infty}$, where the definition of direct sum of Hilbert spaces and related properties are given below.

Under this decomposition, any operator $\mathbf{T} \in \mathfrak{T}(\mathbb{H}_{AB})$ can be regarded as an infinite matrix of operator on $\mathbb{H}_A$ written by

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1n} & \cdots \\ \mathbf{T}_{21} & \mathbf{T}_{22} & \cdots & \mathbf{T}_{2n} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{T}_{n1} & \mathbf{T}_{n2} & \cdots & \mathbf{T}_{nn} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix},$$

where $\mathbf{T}_{ij} \in \mathfrak{T}(\mathbb{H}_A)$. First, suppose that $\mathbf{T}$ is a nonnegative operator. In this case, all the diagonal entries of the above matrix are nonnegative operators on $\mathbb{H}_A$. If the sum $\sum_i \mathbf{T}_{ii}$ converges in the strong operator topology of $\mathfrak{T}(\mathbb{H}_A)$, it is independent of the chosen basis of $\mathbb{H}_B$.

In this case, the partial trace $\mathrm{tr}_B[\mathbf{T}]$ can be computed as follows, where

$$\mathrm{tr}_B[\mathbf{T}] = \mathrm{tr}_B\left[\sum_{i,j} \mathbf{T}_{ij} \otimes \left(|\psi_i\rangle_B\langle\psi_j|\right)\right] = \sum_j \mathbf{T}_{jj}. \tag{2.11}$$

The partial trace of a self-adjoint operator is defined if and only if the partial traces of the positive and negative parts are defined.

We have the following result.

Proposition 2.8.1. *Let $\mathbb{H}_A$ and $\mathbb{H}_B$ be two complex separable Hilbert spaces. Then*

$$\mathrm{tr}_B[\mathbf{R} \otimes \mathbf{S}] = \mathbf{R}\,\mathrm{tr}[\mathbf{S}], \quad \textit{and} \quad \mathrm{tr}_A[\mathbf{R} \otimes \mathbf{S}] = \mathbf{S}\,\mathrm{tr}[\mathbf{R}], \tag{2.12}$$

for all $\mathbf{R} \in \mathfrak{T}(\mathbb{H}_A)$ and for all $\mathbf{S} \in \mathfrak{T}(\mathbb{H}_B)$.

Proof. Let $\{h_i\}_{i=1}^{+\infty}$ be a basis for $\mathbb{H}_A$ and $\{e_j\}_{j=1}^{+\infty}$ be a basis for $\mathbb{H}_B$. For $i, j = 1, 2, \ldots$, let $\mathbf{E}_{ij} : \mathbb{H}_A \to \mathbb{H}_A$ be the map defined by

$$\mathbf{E}_{ij}(h) = \begin{cases} h_j, & \text{if } h = h_i, \\ 0, & \text{otherwise.} \end{cases}$$

For $k, l = 1, 2, \ldots$, let $\mathbf{F}_{kl} : \mathbb{H}_B \to \mathbb{H}_B$ be the map defined by

$$\mathbf{F}_{kl}(e) = \begin{cases} e_l, & \text{if } e = e_k, \\ 0, & \text{otherwise.} \end{cases}$$

Since $\{h_i \otimes e_j \mid i,j = 1,2,\dots\}$ form a basis for $\mathbb{H}_{AB}$, $\{\mathbf{E}_{ij} \otimes \mathbf{F}_{kl} \mid i,j = 1,2,\dots; k,l = 1,2,\dots\}$ form a basis for $\mathfrak{T}(\mathbb{H}_{AB})$. Based on the definition of partial trace (see (2.11)), we have

$$\mathrm{tr}_B[\mathbf{R} \otimes \mathbf{S}] = \mathbf{R}\,\mathrm{tr}[\mathbf{S}], \quad \forall \mathbf{R} \in \mathfrak{T}(\mathbb{H}_A) \text{ and } \forall \mathbf{S} \in \mathfrak{T}(\mathbb{H}_B).$$

The second equality

$$\mathrm{tr}_A[\mathbf{R} \otimes \mathbf{S}] = \mathbf{S}\,\mathrm{tr}[\mathbf{R}], \quad \forall \mathbf{R} \in \mathfrak{T}(\mathbb{H}_A) \text{ and } \forall \mathbf{S} \in \mathfrak{T}(\mathbb{H}_B) \tag{2.13}$$

can be similarly proved. This proves the proposition. □

Proposition 2.8.2. *Let* $\mathbf{S} \in \mathfrak{T}(\mathbb{H}_{AB})$, $\mathbf{T}_A \in \mathfrak{T}(\mathbb{H}_A)$, *and* $\mathbf{I}_B$ *be the identity operator on* $\mathbb{H}_B$. *Then*

$$\mathrm{tr}\big[\mathbf{S}(\mathbf{T}_A \otimes \mathbf{I}_B)\big] = \mathrm{tr}_A[\mathbf{S}_A\mathbf{T}_A] \quad \textit{and} \quad \mathrm{tr}\big[(\mathbf{T}_A \otimes \mathbf{I}_B)\mathbf{S}\big] = \mathrm{tr}_A[\mathbf{T}_A\mathbf{S}_A], \tag{2.14}$$

where $\mathbf{S}_A = \mathrm{tr}_B[\mathbf{S}]$ *is the partial trace of* $\mathbf{S}$ *taken over B.*

Proof. To prove the proposition, we just need the definition of the partial trace: $\mathrm{tr}_B[\mathbf{S}] = \sum_i \langle i|\mathbf{S}|i\rangle_B$, where $\{|i\rangle_B\}_{i=1}^{+\infty}$ is a basis of $\mathbb{H}_B$ and the fact that $\mathrm{tr}[\mathbf{S}] = \mathrm{tr}_A[[\mathrm{tr}_B(\mathbf{S})]]$. The second statement of (2.14) can be proved similarly. This proves the proposition. □

The above statement provides a quantum physics explanation below (see Chang [24]). If $\mathbf{M}$ is an observable on the subsystem A, then the corresponding observable on the composite system is $\mathbf{M} \otimes \mathbf{I}_B$. However, if one chooses to define a reduced state ρ_A, there should be consistency of measurement statistics. The expectation value of $\mathbf{M}$ after the subsystem A is prepared in ρ_A and that of $\mathbf{M} \otimes \mathbf{I}_B$ when the composite system is prepared in ρ should be the same, i. e., the following equality should hold:

$$\mathrm{tr}[\mathbf{M}\rho_A] = \mathrm{tr}\big[(\mathbf{M} \otimes \mathbf{I}_B)\rho\big]. \tag{2.15}$$

This physical interpretation of partial trace will be useful in chapters to follow.

3 Probability measures and convex functions on $\mathcal{S}(\mathbb{H})$

The purpose of this chapter is to investigate probability measures on the set of quantum states $\mathcal{S}(\mathbb{H})$ as a subset of the Banach space of trace-class operators $\mathfrak{T}(\mathbb{H})$. Studies of probability measure and convex functions on $\mathcal{S}(\mathbb{H})$ are essential in computation of Holevo χ-capacity (and hence of classical and quantum capacities) in an unconstrained and constrained quantum channel in later chapters.

Let $\mathbb{H}$ be a complex Hilbert space (finite- or infinite-dimensional) and let $\mathfrak{T}(\mathbb{H})$ be the Banach space of trace-class operators on $\mathbb{H}$ under the trace norm $\|\cdot\|_1$ defined by $\|\mathbf{T}\|_1 = \mathrm{tr}[|\mathbf{T}|]$ for $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$, where $|\mathbf{T}| = \sqrt{\mathbf{T}^*\mathbf{T}}$. Recall that $\mathfrak{T}_+(\mathbb{H}) = \{\rho \in \mathfrak{T}(\mathbb{H}) \mid \rho \geq 0\}$ is the positive cone in $\mathfrak{T}(\mathbb{H})$ and $\mathcal{S}(\mathbb{H}) = \{\rho \in \mathfrak{T}_+(\mathbb{H}) \mid \mathrm{tr}[\rho] = 1\}$ is the space of quantum states on $\mathbb{H}$. We also define $\mathfrak{T}_{\leq 1}(\mathbb{H}) = \{\rho \in \mathfrak{T}_+(\mathbb{H}) \mid \mathrm{tr}[\rho] \leq 1\}$.

We have the following simple observations:

1. $\mathcal{S}(\mathbb{H})$ and $\mathfrak{T}_{\leq 1}(\mathbb{H})$ are closed convex subsets of $\mathfrak{T}(\mathbb{H})$;
2. $\mathcal{S}(\mathbb{H})$ and $\mathfrak{T}_{\leq 1}(\mathbb{H})$ are complete separable metric spaces with the metric defined by the trace norm $\|\cdot\|_1$;
3. $\mathcal{S}(\mathbb{H})$ is compact in $\mathfrak{T}_+(\mathbb{H})$ if $\dim(\mathbb{H}) < +\infty$. However, $\mathcal{S}(\mathbb{H})$ is not compact in general but only weakly compact in $\mathfrak{T}_+(\mathbb{H})$ if $\dim(\mathbb{H}) = +\infty$.

3.1 Probability measures on $\mathcal{S}(\mathbb{H})$

3.1.1 Support of Borel measures

Consider the Borel measurable space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$, where $\mathbb{X}$ is a locally compact Hausdorff space (for the purpose of this book, we can simply assume that $\mathbb{X}$ is either a separable Banach or Hilbert space or a closed subset of a Banach or Hilbert space) and $\mathcal{B}(\mathbb{X})$ is the Borel σ-algebra of open subsets of $\mathbb{X}$.

We also recall some basic definitions in measure theory (see, e. g., Halmos [58] and Rudin [133]) as follows.

If $\mu : \mathcal{B}(\mathbb{X}) \to [0, +\infty]$ is a function such that (i) $\mu(B) \geq 0$ for all $B \in \mathcal{B}(\mathbb{X})$; (ii) $\mu(\emptyset) = 0$ and (iii) $\mu(\bigcup_{i=1}^{+\infty} A_i) = \sum_{i=1}^{+\infty} \mu(A_i)$ for any sequence $(A_i)_{i=1}^{+\infty}$ of pairwise disjoint Borel sets (i. e., $A_i \cap A_j = \emptyset$ if $i \neq j$), then μ is called a Borel measure on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. In this case, the triplet $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ will be referred to as a Borel measure space.

A Borel measure μ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ is said to be inner regular if

$$\mu(B) = \sup\{\mu(F) \mid \text{ compact } F \subseteq B\}, \quad \forall B \in \mathcal{B}(\mathbb{X}),$$

is said to be outer regular if

$$\mu(B) = \inf\{\mu(G) \mid \text{ open } G \supseteq B\}, \quad \forall B \in \mathcal{B}(\mathbb{X}).$$

https://doi.org/10.1515/9783110788105-003

The Borel measure μ is regular if it is both inner regular and outer regular. The Borel measure μ is said to be a Radon measure if it is both inner regular and locally finite (i. e., $\mu(F) < +\infty$ for all compact Borel set F). If the Borel measure μ is such that $\mu(\mathbb{X}) = 1$, then μ is called a Borel probability measure and the triplet $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ is called a Borel probability space. The collection of all Borel probability measures on $\mathbb{X}$ will be denoted by $\mathcal{P}(\mathbb{X})$.

Without further mention, all Borel measures discussed in this book are assumed to be regular Borel measures.

Definition 3.1.1. Let μ be a Borel measure on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. Then $\text{supp}(\mu)$, the support (or spectrum) of μ, is defined to be the set of all points x in $\mathbb{X}$ for which every open neighborhood N_x of x has positive measure. In other words,

$$\text{supp}(\mu) = \{x \in \mathbb{X} \mid \mu(N_x) > 0 \text{ for all open set } N_x \text{ containing } x\}.$$

We have the following results regarding $\text{supp}(\mu)$.

Proposition 3.1.2. *Let* $(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ *be a Borel measure space. Then* $\text{supp}(\mu)$ *has the following properties:*
1. *A measure μ on* $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ *is strictly positive (i. e., $\mu(B) > 0$ for any nonempty open subset $B \subset \mathbb{X}$) if and only if* $\text{supp}(\mu) = \mathbb{X}$.
2. $\text{supp}(\mu)$ *is a closed subset of* $\mathbb{X}$.

Proof. 1. ($\Rightarrow$) It is clear that $\text{supp}(\mu) \subseteq \mathbb{X}$. If μ is strictly positive and if $x \in \mathbb{X}$ is arbitrary, then any open neighborhood of x has positive measure, since it is a nonempty open set. Hence, $x \in \text{supp}(\mu)$, this shows that $\mathbb{X} \subseteq \text{supp}(\mu)$. Therefore, $\text{supp}(\mu) = \mathbb{X}$.

($\Leftarrow$) Conversely, assume that $\text{supp}(\mu) = \mathbb{X}$. Then every nonempty open set (being an open neighborhood of some point in its interior, which is also a point of the support) has positive measure. Hence, μ is strictly positive.

2. $\text{supp}(\mu)$ is closed in $\mathbb{X}$, because the complement of $\text{supp}(\mathbb{X})$ is the union of the open sets of measure 0. That is, $(\text{supp}(\mathbb{X}))^c$, the complement of $\text{supp}(\mu)$ in $\mathbb{X}$, is open. Therefore, $\text{supp}(\mathbb{X})$ is closed. This proves the proposition. □

For the remainder of this section, we concentrate our attention on the case $\mathbb{X} = \mathcal{S}(\mathbb{H})$, where $\mathcal{S}(\mathbb{H})$ is the collection of quantum states (i. e., all positive bounded linear operators with unit trace) on the Hilbert space $\mathbb{H}$ and also when $\mathbb{X} = \mathcal{P}(\mathcal{S}(\mathbb{H}))$, the space of Borel probability measures on $\mathcal{S}(\mathbb{H})$.

A (Borel) probability measure μ on the Borel σ-algebra of subsets of $\mathcal{S}(\mathbb{H})$ is said to be an atomic (or discrete) measure if its support $\text{supp}(\mu)$ consists of countably infinite or finite number of elements $\{\rho_i\}_{i=1}^N$, where $N \leq +\infty$, in $\mathcal{S}(\mathbb{H})$. In this case, each of these elements ρ_i is called an atom in the Borel probability space $(\mathcal{S}(\mathbb{H}), \mathcal{B}(\mathcal{S}(\mathbb{H})), \mu)$. Denote $\mathcal{P}^{\text{dis}}(\mathcal{S}(\mathbb{H}))$ be the subset of $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ consisting of atomic (or discrete) measures.

We often use the following terminologies.

(A) Discrete ensembles

For convenience, an atomic (or discrete) measure $\mu \in \mathcal{P}^{\mathrm{dis}}(\mathcal{S}(\mathbb{H}))$ can normally be represented as $\mu = \{p_i, \rho_i\}$, where $\rho_i \in \mathcal{S}(\mathbb{H})$ ($\rho_i \neq \rho_j$ for $i \neq j$) denotes an atom and $p_i > 0$ denotes the probability mass assigned to the atom ρ_i, where $\sum_i p_i = 1$. In this case, $\{p_i, \rho_i\}$ will be referred to as a *discrete ensemble*.

As described in Oreshkov and Calsamiglia [120], the concept of ensemble of states can be used to describe situations in which a system takes a state ρ_i at random with probability p_i. The statement that a system takes the state ρ_x means that there exists classical information x about the identity of the state. This is to be distinguished from the situation in which no information about the identity of the state exists or can be obtained. In the latter case, for all practical purposes, the average density operator of the ensemble, $\rho = \sum_i p_i \rho_i$, provides a complete description of the state of the system. An example of an ensemble of states is the output of a nondestructive generalized measurement (see Linblad [106], Bratteli and Robinson [15] and Shirokov [150]). Under the most general type of *quantum measurement*, a density operator ρ transforms as

$$\rho \mapsto \rho_i = \frac{\mathfrak{M}_i(\rho)}{\mathrm{tr}[\mathfrak{M}_i(\rho)]} \quad \text{with probability } p_i = \mathrm{tr}[\mathfrak{M}_i(\rho)], \tag{3.1}$$

where $\mathfrak{M}_i(\cdot) = \sum_j \mathbf{M}_{ij}(\cdot)\mathbf{M}_{ij}^*$ is the measurement superoperator corresponding to outcome i. (The operators $\mathbf{M}_{ij} : \mathcal{S}(\mathbb{H}) \to \mathcal{S}(\mathbb{H})$ satisfy the completeness relation $\sum_{i,j} \mathbf{M}_{ij}^* \mathbf{M}_{ij} = \mathbf{I}$.)

(B) Continuous ensembles

A Borel probability measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$, which is not a discrete ensemble (atomic measure), will be referred to as a *continuous ensemble*. The subset of $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ that consists of continuous ensembles will be denoted by $\mathcal{P}^{\mathrm{con}}(\mathcal{S}(\mathbb{H}))$. Therefore, the space $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ can be decomposed as

$$\mathcal{P}(\mathcal{S}(\mathbb{H})) = \mathcal{P}^{\mathrm{dis}}(\mathcal{S}(\mathbb{H})) \cup \mathcal{P}^{\mathrm{con}}(\mathcal{S}(\mathbb{H})).$$

The concept of support of a Borel measure (see Definition 3.1.1) and that of spectrum (see Definition 1.4.2) of a self-adjoint linear operator on a Hilbert space are closely related (see Proposition 1.4.3).

Concepts of discrete ensemble and continuous ensembles in quantum states was first introduced in Oreshkov and Calsamiglia [120] and applied to the context of infinite-dimensional quantum information by Holevo and Shirokov [81].

Example 3.1. To explore the relation between the support of a probability measure and the spectrum of a self-adjoint linear operator on a Hilbert space, let μ be a regular Borel measure on the Borel measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Consider the multiplication operator $(\mathbf{A}f)(x) = xf(x)$ on its natural domain

$$\text{dom}(\mathbf{A}) = \{f \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu) \mid xf(x) \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)\}.$$

The multiplication operator $\mathbf{A}$ defined is a self-adjoint operator on the complex Hilbert space $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ equipped with the inner product $\langle \cdot, \cdot \rangle_{L^2}$ (see Example 1.3). Indeed,

$$\begin{aligned}\langle f, \mathbf{A}g\rangle_{L^2} &= \int_{\mathbb{R}} \overline{f(x)} x g(x) \mu(dx) \\ &= \int_{\mathbb{R}} x\overline{f(x)} g(x) \mu(dx) \\ &= \langle \mathbf{A}f, g\rangle_{L^2}, \quad \forall f, g \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu).\end{aligned}$$

Thus, $\mathbf{A}$ is a self-adjoint linear operator on $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ and its spectrum coincides with the essential range of the identity function, which is precisely the support of μ.

3.2 Some compactness criteria

Let $\mathbb{X}$ be a complex Banach or Hilbert space equipped with Banach or Hilbertian norm $\|\cdot\|_{\mathbb{X}}$. Recall our discussion of topological properties of $\mathbb{X}$ from Section 1.1 that a closed subset $\mathcal{K} \subset \mathbb{X}$ is compact if every open covering of $\mathcal{K}$ has finite subcovering. That is, if $\mathcal{K} \subset \bigcup_{\lambda \in \Lambda} \mathcal{O}_\lambda$, where $\{\mathcal{O}_\lambda, \lambda \in \Lambda\}$ is a family of open sets in $\mathbb{X}$, then there exists a finite subfamily $\{\mathcal{O}_{\lambda_i}, i = 1, 2, \ldots, N\}$, where $\lambda_i \in \Lambda$ for $i = 1, 2, \ldots, N$, such that $\mathcal{K} \subset \bigcup_{i=1}^N \mathcal{O}_{\lambda_i}$. Equivalently, the compactness of $\mathcal{K}$ in $\mathbb{X}$ can also be characterized as follows. Every sequence $(\varphi_n)_{n=1}^{+\infty}$ in $\mathcal{K}$ has a subsequence $(\varphi_{n_k})_{k=1}^{+\infty}$ that converges to some $\varphi \in \mathcal{K}$ under the norm $\|\cdot\|_{\mathbb{X}}$.

In the following two subsections, we give characterization of compactness subset of $\mathcal{S}(\mathbb{H})$ and of $\mathcal{P}(\mathcal{S}(\mathbb{H}))$, respectively. The presentation of this section is based on the results in Shirokov [144–146] and Protasov and Shirokov [127].

3.2.1 Compactness criteria on $\mathcal{S}(\mathbb{H})$

We have the following compactness criterion, due originally to Holevo and Shirokov [81] for closed subsets of $\mathcal{S}(\mathbb{H})$ (see Holevo and Shirokov [81] and Chang [22, 24]).

Theorem 3.2.1 (Prohorov's compactness criterion). *A closed subset $\mathcal{K} \subset \mathcal{S}(\mathbb{H})$ is compact (in the trace-class norm) if and only if for every $\epsilon > 0$ there is a finite-dimensional projection $\mathbf{P}_\epsilon$ on $\mathbb{H}$ such that $\text{tr}[\rho \mathbf{P}_\epsilon] \geq 1 - \epsilon$ for all $\rho \in \mathcal{K}$.*

Proof. ($\Rightarrow$). We follow the proof provided in Chang [22] for necessity condition as follows. Let $\mathcal{K}$ be a compact subset of $\mathcal{S}(\mathbb{H})$. We want to show that for every $\epsilon > 0$ there is a finite-dimensional projection $\mathbf{P}_\epsilon$ on $\mathbb{H}$ such that $\text{tr}[\rho \mathbf{P}_\epsilon] \geq 1 - \epsilon$ for all $\rho \in \mathcal{K}$. Suppose this were not true for contradiction purpose. Then there is an $\epsilon > 0$ such that for any arbitrary finite rank projection operator $\mathbf{P}$ there exists a state $\rho \in \mathcal{K}$ such that $\mathrm{tr}[\rho\mathbf{P}] < 1-\epsilon$. Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be a sequence of finite rank projections on $\mathbb{H}$ monotonically converging to the identity operator $\mathbf{I}_{\mathbb{H}}$ on $\mathbb{H}$ in the weak operator topology. That is,

$$\lim_{n\to+\infty}\langle\phi,\mathbf{P}_n\psi\rangle_{\mathbb{H}} = \langle\phi,\mathbf{I}_{\mathbb{H}}\psi\rangle_{\mathbb{H}} = \langle\phi,\psi\rangle_{\mathbb{H}}, \quad \forall\phi,\psi\in\mathbb{H}.$$

Let $(\rho_n)_{n=1}^{+\infty}$ be the corresponding sequence of states in $\mathcal{K}$ such that $\mathrm{tr}[\rho_n\mathbf{P}_n] < 1-\epsilon$ for all $n \geq 1$. Since $\mathcal{K}$ is compact, there exists a subsequence $(\rho_{n_k})_{k=1}^{+\infty}$ of $(\rho_n)_{n=1}^{+\infty}$ converging to a state $\rho_* \in \mathcal{K}$. Therefore, by construction, we have

$$\mathrm{tr}[\rho_{n_k}\mathbf{P}_{n_l}] \leq \mathrm{tr}[\rho_{n_k}\mathbf{P}_{n_k}] < 1-\epsilon, \quad \forall k > l.$$

Hence,

$$\mathrm{tr}[\rho_*] = \mathrm{tr}[\rho_*\mathbf{I}_{\mathbb{H}}] = \lim_{l\to+\infty}\mathrm{tr}[\rho_*\mathbf{P}_{n_l}] = \lim_{l\to+\infty}\lim_{k\to+\infty}\mathrm{tr}[\rho_{n_k}\mathbf{P}_{n_l}] < 1-\epsilon.$$

This contradict the fact that $\rho_* \in \mathcal{K} \subset \mathcal{S}(\mathbb{H})$ since all elements in $\mathcal{S}(\mathbb{H})$ have unit trace.

($\Leftarrow$). We follow the proof provided by Holevo and Shirokov [81] for sufficiency condition as follows. Suppose $\mathcal{K}$ is a subset of $\mathcal{S}(\mathbb{H})$ satisfying the compactness criterion. Let $(\rho_n)_{n=1}^{+\infty}$ be an arbitrary sequence in $\mathcal{K}$. Since the closed unit ball $B(\mathbf{0};1) := \{\mathbf{A} \in \mathfrak{B}(\mathbb{H}) \mid \|\mathbf{A}\|_\infty \leq 1\}$ is compact in the weak operator (see the Banach–Alaoglu Theorem 1.1.4), there exists a subsequence $(\rho_{n_k})_{k=1}^{+\infty}$ of $(\rho_n)_{n=1}^{+\infty}$ converging to a positive operator ρ_* in the weak topology or equivalently in the trace norm $\|\cdot\|_1$. We have

$$\mathrm{tr}[\rho_*] \leq \liminf_{k\to+\infty}\mathrm{tr}[\rho_{n_k}] = 1.$$

Therefore, to prove that ρ_* is a state it is sufficient to show that $\mathrm{tr}[\rho_*] \geq 1$. Let $\epsilon > 0$ and $\mathbf{P}_\epsilon$ be the corresponding projector. Then

$$\mathrm{tr}[\rho_*] = \mathrm{tr}[\rho_*\mathbf{I}_{\mathbb{H}}] \geq \mathrm{tr}[\rho_*\mathbf{P}_\epsilon] = \lim_{k\to+\infty}\mathrm{tr}[\rho_{n_k}\mathbf{P}_\epsilon] \geq 1-\epsilon,$$

where the equality follows from the finite dimensionality of the space $\mathbf{P}_\epsilon(\mathbb{H})$. Thus, ρ_* is a state. The proof given above implies that the subsequence $(\rho_{n_k})_{k=1}^{+\infty}$ converges to the state ρ_* in the trace norm $\|\cdot\|_1$. This shows that $\mathcal{K}$ is compact in trace norm. This proves the theorem. □

Recall the notation that $\mathfrak{T}_{\leq 1}(\mathbb{H}) = \{\rho \in \mathfrak{T}(\mathbb{H}) \mid \rho \geq 0, \mathrm{tr}[\rho] \leq 1\}$.

The following compactness criterion for subsets of $\mathfrak{T}_{\leq 1}(\mathbb{H})$ can be proved by a simple modification of arguments used in the proof of Theorem 3.2.1.

Proposition 3.2.2. *A closed subset $\mathcal{A}$ of $\mathfrak{T}_{\leq 1}(\mathbb{H})$ is compact if and only if for an arbitrary $\epsilon > 0$ there exists a finite rank projector $\mathbf{P}_\epsilon$ such that $\mathrm{tr}[(\mathbf{I}_{\mathbb{H}} - \mathbf{P}_\epsilon)\mathbf{A}] < \epsilon$ for all $\mathbf{A} \in \mathcal{A}$.*

Consider quantum systems A and B represented by Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. In the following, we explore compactness criteria of subsets in $\mathcal{S}(\mathbb{H}_{AB})$ as well as of $\mathfrak{T}_{\leq 1}(\mathbb{H}_{AB})$. Recall that $\mathcal{S}(\mathbb{H}_{AB})$ is the space of quantum states on the composite quantum system $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$ defined by

$$\mathcal{S}(\mathbb{H}_{AB}) = \{\upsilon \in \mathfrak{T}(\mathbb{H}_{AB}) \mid \upsilon \geq 0, \mathrm{tr}[\upsilon] = 1\}$$

and

$$\mathfrak{T}_{\leq 1}(\mathbb{H}_{AB}) = \{\upsilon \in \mathfrak{T}(\mathbb{H}_{AB}) \mid \upsilon \geq 0, \mathrm{tr}[\upsilon] \leq 1\}.$$

Let $\mathcal{A}$ and $\mathcal{B}$ be subsets of $\mathfrak{T}_{\leq 1}(\mathbb{H}_A)$ and $\mathfrak{T}_{\leq 1}(\mathbb{H}_B)$, respectively. Let $\mathcal{A} \otimes \mathcal{B}$ be the subset of $\mathfrak{T}_{\leq 1}(\mathbb{H}_{AB})$ defined by

$$\mathcal{A} \otimes \mathcal{B} = \{\mathbf{C} \in \mathfrak{T}_{\leq 1}(\mathbb{H}_{AB}) \mid \mathrm{tr}_B[\mathbf{C}] \in \mathcal{A} \text{ and } \mathrm{tr}_A[\mathbf{C}] \in \mathcal{B}\}.$$

The following corollary can be found in Holevo and Shirokov [81].

Corollary 3.2.3. *$\mathcal{A} \otimes \mathcal{B}$ is a compact subset of $\mathfrak{T}_{\leq 1}(\mathbb{H}_A) \otimes \mathfrak{T}_{\leq 1}(\mathbb{H}_B)$ if and only if $\mathcal{A}$ and $\mathcal{B}$ are compact subsets of $\mathfrak{T}_{\leq 1}(\mathbb{H}_A)$ and $\mathfrak{T}_{\leq 1}(\mathbb{H}_B)$, respectively.*

Proof. ($\Rightarrow$). The compactness of the set $\mathcal{A} \otimes \mathcal{B}$ implies the compactness of the sets $\mathcal{A}$ and $\mathcal{B}$ due to continuity of the partial trace.

($\Leftarrow$). Let $\mathcal{A}$ and $\mathcal{B}$ be compact. By Proposition 3.2.2, for an arbitrary $\epsilon > 0$ there exist finite rank projectors $\mathbf{P}_\epsilon$ and $\mathbf{Q}_\epsilon$ such that

$$\mathrm{tr}[\mathbf{P}_\epsilon \mathbf{A}] > \mathrm{tr}[\mathbf{A}] - \epsilon, \quad \forall \mathbf{A} \in \mathcal{A}, \text{ and } \mathrm{tr}[\mathbf{Q}_\epsilon \mathbf{B}] > \mathrm{tr}[\mathbf{B}] - \epsilon, \ \forall \mathbf{B} \in \mathcal{B}.$$

Since $\mathbf{C}_A = \mathrm{tr}_B[\mathbf{C}] \in \mathcal{A}$ and $\mathbf{C}_B = \mathrm{tr}_A[\mathbf{C}] \in \mathcal{B}$ for an arbitrary $\mathbf{C} \in \mathcal{A} \otimes \mathcal{B}$, we have

$$\begin{aligned} &\mathrm{tr}[(\mathbf{P}_\epsilon \otimes \mathbf{Q}_\epsilon)\mathbf{C}] \\ &\quad = \mathrm{tr}[(\mathbf{P}_\epsilon \otimes \mathbf{I}_B)\mathbf{C}] - \mathrm{tr}[(\mathbf{P}_\epsilon \otimes (\mathbf{I}_B - \mathbf{Q}_\epsilon))\mathbf{C}] \\ &\quad \geq \mathrm{tr}[\mathbf{P}_\epsilon \mathbf{C}_A] - \mathrm{tr}[(\mathbf{I}_B - \mathbf{Q}_\epsilon)\mathbf{C}_B] > \mathrm{tr}[\mathbf{C}] - 2\epsilon, \end{aligned}$$

where $\mathbf{I}_B := \mathbf{I}_{\mathbb{H}_B}$ is the identity operator on $\mathbb{H}_B$. Proposition 3.2.2 implies the compactness of the set $\mathcal{A} \otimes \mathcal{B}$. This proves the corollary. □

Definition 3.2.4. An unbounded positive linear operator $\mathbf{H}$ on $\mathbb{H}$ is said to be an $\mathfrak{H}$-operator if it has a discrete point spectrum (eigenvalues),

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \lambda_n \leq \cdots$$

Let $\mathbf{Q}_n$ be the spectral projector of $\mathbf{H}$ corresponding to the lowest n eigenvalues. That is,

$$\mathbf{Q}_n(\phi) = \sum_{i=1}^{n} \lambda_i \langle \phi, \varphi_i \rangle_{\mathbb{H}} \varphi_i, \quad \forall \phi \in \mathbb{H},$$

where $\varphi_i \in \mathbb{H}$ is the eigenvector corresponding to λ_i for $i = 1, 2, \ldots$.

We have the following observations/comments on $\mathfrak{H}$-operators:

1. An $\mathfrak{H}$-operator $\mathbf{H}$ can be interpreted as a quantum observable (or Hamiltonian) on the quantum system $\mathbb{H}$ with energy levels $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \le \cdots$.
2. If $\{\varphi_i\}_{i=1}^{+\infty}$ is an orthonormal set of eigenvectors corresponding to the collection of eigenvalues $\{\lambda_n\}_{n=1}^{+\infty}$ of the $\mathfrak{H}$-operator $\mathbf{H}$. Then $\mathbf{H}$ has the spectral representation

$$\mathbf{H} = \sum_{i=1}^{+\infty} \lambda_i |\varphi_i\rangle_{\mathbb{H}} \langle \varphi_i|$$

   on the domain of $\mathbf{H}$, dom($\mathbf{H}$), defined by

$$\operatorname{dom}(\mathbf{H}) = \left\{ \varphi \in \overline{\operatorname{span}}(\{|\varphi_i\rangle_{\mathbb{H}}\langle\varphi_i|\}_{i=1}^{+\infty}) \mid \sum_{i=1}^{+\infty} \lambda_i^2 \langle \varphi, \varphi_i \rangle_{\mathbb{H}} < +\infty \right\},$$

   where $\overline{\operatorname{span}}(\ldots)$ is the closure under $\|\cdot\|_{\mathbb{H}}$ of the span of the set $(\cdots)$.
3. Let $(\mathbf{Q}_n)_{n=1}^{+\infty}$ be the increasing sequence of spectral projections for an $\mathfrak{H}$-operator $\mathbf{H}$ on the Hilbert space $\mathbb{H}$, then $(\mathbf{Q}_n)_{n=1}^{+\infty}$ converges to the identity operator $\mathbf{I}_{\mathbb{H}}$ under the operator norm $\|\cdot\|_\infty$.
4. Following Holevo [72, 73], we define

$$\operatorname{tr}[\rho \mathbf{H}] = \lim_{n\to+\infty} \operatorname{tr}[\rho \mathbf{Q}_n \mathbf{H}], \quad \forall \rho \in \mathcal{S}(\mathbb{H}), \tag{3.2}$$

   where the sequence on the right-hand side of the above equation is monotonously increasing.

Example 3.2. An important example of an $\mathfrak{H}$-operator is $\mathbf{H} = -\log \mathbf{A}$, where $\mathbf{A}$ is any operator that is represented as $\mathbf{A} = \sum_{i=1}^{+\infty} \lambda_i |\varphi_i\rangle_{\mathbb{H}}\langle\varphi_i|$ in $\mathfrak{T}_{\le 1}(\mathbb{H})$ (where $\mathfrak{T}_{\le 1}(\mathbb{H}) = \{\mathbf{A} \in \mathfrak{T}_+(\mathbb{H}) \mid \operatorname{tr}[\mathbf{A}] \le 1\}$) with infinite sequence of positive eigenvalues ($\lambda_i > 0$, $i = 1, 2, \ldots$), which has the representation

$$\mathbf{H} = -\log \mathbf{A} = \sum_{i=1}^{+\infty} (-\log \lambda_i) |\varphi_i\rangle_{\mathbb{H}} \langle \varphi_i|$$

on the domain

$$\operatorname{dom}(-\log \mathbf{A}) = \left\{ \phi \in \operatorname{supp}(\mathbf{A}) \,\middle|\, \sum_{i=1}^{+\infty} (\log \lambda_i)^2 |\langle \varphi_i, \phi \rangle_{\mathbb{H}}|^2 < +\infty \right\}.$$

Note that for any operators $\mathbf{A} \in \mathfrak{T}_1(\mathbb{H})$ and $\mathbf{B} \in \mathfrak{T}_1(\mathbb{K})$ the following identity holds:

$$-\log(\mathbf{A} \otimes \mathbf{B}) = (-\log \mathbf{A}) \otimes \mathbf{I}_{\mathbb{K}} + \mathbf{I}_{\mathbb{H}} \otimes (-\log \mathbf{B}),$$

where "=" means the operators coincides on

$$\operatorname{dom}(-\log(\mathbf{A} \otimes \mathbf{B})) \subseteq \operatorname{supp}(\mathbf{A}) \otimes \operatorname{supp}(\mathbf{B}).$$

Let $\mathcal{K}_{\mathbf{H}}(h) \subset \mathcal{S}(\mathbb{H})$, the set of quantum states under which the observable $\mathbf{H}$ has expected values less or equal to h, be defined by

$$\mathcal{K}_{\mathbf{H}}(h) = \{\rho \in \mathcal{S}(\mathbb{H}) \mid \operatorname{tr}[\rho \mathbf{H}] \le h\}, \quad h \in \mathbb{R}_+, \tag{3.3}$$

where $\mathbf{H}$ is an $\mathfrak{H}$-operator.

The following result is due to Holevo [72].

Theorem 3.2.5 (Characterization of compact subsets of $\mathcal{S}(\mathbb{H})$). *$\mathcal{K} \subset \mathcal{S}(\mathbb{H})$ is compact under trace norm $\|\cdot\|_1$ if and only if there exist an $\mathfrak{H}$-operator $\mathbf{H}$ and an $h \ge 0$ such that*

$$\mathcal{K} = \{\rho \in \mathcal{S}(\mathbb{H}) \mid \operatorname{tr}[\rho \mathbf{H}] \le h\}. \tag{3.4}$$

Proof. ($\Rightarrow$) Assume that $\mathcal{K} \subset \mathcal{S}(\mathbb{H})$ is compact under the trace norm $\|\cdot\|_1$. Since $\mathcal{K} \subset \mathcal{S}(\mathbb{H})$ is compact, Theorem 3.2.1 implies that for any natural number n there exists a finite rank projector $\mathbf{P}_n$ such that $\operatorname{tr}[\rho \mathbf{P}_n] \ge 1 - n^{-3}$ for all $\rho \in \mathcal{K}$. Without loss of generality, we may assume that $\bigvee_{k=1}^{+\infty} \mathbf{P}_k(\mathbb{H}) = \mathbb{H}$, where $\bigvee$ denotes closed linear span of subspaces. Let $\hat{\mathbf{P}}_n$ be the projector on the finite-dimensional subspace $\bigvee_{k=1}^{n} \mathbf{P}_k(\mathbb{H})$. Thus, $\mathbf{H} = \sum_{n=1}^{+\infty} n(\hat{\mathbf{P}}_{n+1} - \hat{\mathbf{P}}_n)$ is an $\mathfrak{H}$-operator satisfying

$$\operatorname{tr}[\rho \mathbf{H}] = \sum_{n=1}^{+\infty} n \operatorname{tr}[\rho(\hat{\mathbf{P}}_{n+1} - \hat{\mathbf{P}}_n)] \le \sum_{n=1}^{+\infty} n \operatorname{tr}[\rho(\mathbf{I}_{\mathbb{H}} - \hat{\mathbf{P}}_n)] \le \sum_{n=1}^{+\infty} n^{-2} = h$$

for arbitrary $\rho \in \mathcal{K}$.

($\Leftarrow$) Without loss of generality, we assume that the eigenvalues $\{\lambda_n\}_{n=1}^{+\infty}$ of the $\mathfrak{H}$-operator $\mathbf{H}$ is monotonously increasing and denote by $\mathbf{P}_n$ the finite-dimensional projection onto the eigenspace corresponding to the first n eigenvalues, then $\mathbf{P}_n \uparrow \mathbf{I}_{\mathbb{H}}$. By a general criterion, a weakly closed subset $\mathcal{S}$ of density operators is weakly compact if and only if for every $\epsilon > 0$ there is a finite-dimensional projection $\mathbf{P}$ such that $\operatorname{tr}[\mathbf{S}(\mathbf{I}_{\mathbb{H}} - \mathbf{P})] \le \epsilon$ for all $\mathbf{S} \in \mathcal{S} \subset \mathcal{S}(\mathbb{H})$. However, the weak convergence of density operators is equivalent to their trace-norm convergence. Since $\lambda_{n+1}(\mathbf{I}_{\mathbb{H}} - \mathbf{P}_n) \le \mathbf{H}$, we have $\operatorname{tr}[\mathbf{S}(\mathbf{I}_{\mathbb{H}} - \mathbf{P}_n)] \le \lambda_{n+1}^{-1} \operatorname{tr}[\mathbf{S}\mathbf{H}] \le \lambda_{n+1}^{-1} h$ for $\mathbf{S} \in \mathcal{K}_{\mathbf{H}}(h)$. This shows that the set $\mathcal{K}_{\mathbf{H}}(h)$ is a compact subset of $\mathcal{S}(\mathbb{H})$ for each $h > 0$. This proves the theorem. □

3.2.2 Compactness criteria on $\mathcal{P}(\mathcal{S}(\mathbb{H}))$

Let $\mathcal{A}$ be a closed subset of $\mathcal{S}(\mathbb{H})$ under trace-norm $\|\cdot\|_1$. We equip $\mathcal{P}(\mathcal{A})$, the space of Borel probability measures on $\mathcal{A}$, with the topology of weak convergence (see Billingsley [11] and Parthasarathy [122]). Recall that a sequence of Borel probability measures $(\mu_n)_{n=1}^{+\infty} \subset \mathcal{P}(\mathcal{A})$ is said to converge to $\mu \in \mathcal{P}(\mathcal{A})$ weakly if

$$\lim_{n\to\infty} \int_{\mathcal{A}} f(\rho)\mu_n(d\rho) = \int_{\mathcal{A}} f(\rho)\mu(d\rho), \quad \forall f \in C_b(\mathcal{A}), \tag{3.5}$$

where $C_b(\mathcal{A})$ is the space of bounded continuous real-valued functions on $\mathcal{A}$.

Since $\mathcal{A}$ is a closed subset of $\mathcal{S}(\mathbb{H})$, it is a complete separable metric space under the trace-norm $\|\cdot\|_1$. In this case, $\mathcal{P}(\mathcal{A})$ can be considered as a complete separable metric space under the topology of weak convergence (see Parthasarathy [122]).

Definition 3.2.6. A set $\mathcal{P}$ of Borel probability measures $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ on $\mathcal{S}(\mathbb{H})$ is said to be tight if, for each $\epsilon > 0$, there is a compact set $\mathcal{K} \subset \mathcal{S}(\mathbb{H})$ such that

$$\mu(\mathcal{K}) > 1 - \epsilon, \quad \forall \mu \in \mathcal{P}. \tag{3.6}$$

The proof of the following Prokhorov theorem can be found in Billingsley [11] or Parsatherathy [122] for any general metric space.

Theorem 3.2.7 (Prokhorov theorem). *A subset $\mathcal{P}$ of $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ is tight if and only if it is relatively compact under weak convergence.*

In the following, we explore compactness properties of probability measures defined on a closed subset $\mathcal{K}$ of $\mathcal{S}(\mathbb{H})$.

Proposition 3.2.8. *The set $\mathcal{P}(\mathcal{K})$ is a compact subset of $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ if and only if the set $\mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{H})$.*

Proof. ($\Rightarrow$) Assume that the set $\mathcal{P}(\mathcal{K})$ is a compact subset of $\mathcal{P}(\mathcal{S}(\mathbb{H}))$. The set $\mathcal{K}$ is the image of the set $\mathcal{P}(\mathcal{K})$ under the continuous mapping $\mu \mapsto \bar{\rho}(\mu)$ (see Proposition 3.3.5 in the next section); hence, it is compact.

($\Leftarrow$) Assume that $\mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{H})$. By Theorem 3.2.5, there exists an $\mathfrak{H}$-operator $\mathbf{H}$ and a positive $h > 0$ such that $\mathrm{tr}[\rho\mathbf{H}] \leq h$ for all $\rho \in \mathcal{K}$. For arbitrary $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$, we have

$$\begin{aligned}
&\int_{\mathcal{S}(\mathbb{H})} \mathrm{tr}[\rho\mathbf{H}]\mu(d\rho) \\
&= \int_{\mathcal{S}(\mathbb{H})} \Big(\lim_{n\to+\infty} \mathrm{tr}[\rho\mathbf{Q}_n\mathbf{H}]\Big)\mu(d\rho) \quad \text{(by (3.2))} \\
&= \lim_{n\to+\infty}\Big(\int_{\mathcal{S}(\mathbb{H})} \mathrm{tr}[\rho\mathbf{Q}_n\mathbf{H}]\mu(d\rho)\Big)
\end{aligned}$$

$$\text{(by the monotone convergence theorem)}$$
$$= \lim_{n\to+\infty} \operatorname{tr}\Big[\int_{\mathcal{S}(\mathbb{H})} (\rho\mathbf{Q}_n\mathbf{H})\mu(d\rho)\Big]$$
$$\text{(by linearity of trace of the finite-dimensional operator)}$$
$$= \operatorname{tr}\Big[\lim_{n\to+\infty} \int_{\mathcal{S}(\mathbb{H})} (\rho\mathbf{Q}_n\mathbf{H})\mu(d\rho)\Big]$$
$$\text{(by continuity of the finite-dimensional trace-class operator)}$$
$$= \operatorname{tr}\Big[\int_{\mathcal{S}(\mathbb{H})} \lim_{n\to+\infty} (\rho\mathbf{Q}_n\mathbf{H})\mu(d\rho)\Big]$$
$$\text{(by the monotone convergence theorem)}$$
$$= \operatorname{tr}\Big[\Big(\int_{\mathcal{S}(\mathbb{H})} \rho\mu(d\mu)\Big)\mathbf{H}\Big]$$
$$= \operatorname{tr}[\overline{\rho}(\mu)\mathbf{H}] \leq h.$$

Let $\mathcal{K}_\epsilon = \{\rho \in \mathcal{S}(\mathbb{H}) \mid \operatorname{tr}[\rho\mathbf{H}] \leq h\epsilon^{-1}\}$. The set $\mathcal{K}_\epsilon$ is a compact subset of $\mathcal{S}(\mathbb{H})$ for any ϵ. For any measure μ in $\mathcal{P}(\mathcal{K}_\epsilon)$, we therefore have

$$\mu(\mathcal{S}(\mathbb{H})\backslash\mathcal{K}_\epsilon) = \int_{\mathcal{S}(\mathbb{H})\backslash\mathcal{K}_\epsilon} \mu(d\rho) \leq \epsilon h^{-1} \int_{\mathcal{S}(\mathbb{H})\backslash\mathcal{K}_\epsilon} (\operatorname{tr}[\rho\mathbf{H}])\mu(d\rho) \leq \epsilon. \tag{3.7}$$

By the compactness criterion (Theorem 3.2.1), the closed set $\mathcal{P}(\mathcal{K})$ is compact. This proves the proposition. □

3.3 Barycenters

Concept of barycenter plays an important role in infinite-dimensional quantum information theory. The presentation of materials in this section is largely based on the results obtained by Holevo and Shriokov [81], Shirokov [146] and Shirokov and Holevo [158].

3.3.1 A brief review of convex analysis

We recall some facts about the convex analysis that are relevant to development of quantum information theory below. Readers are referred to Rockafellar [131] for detailed accounts on convex analysis.

Let $\mathcal{A}$ be a subset of $\mathbb{X}$, where $\mathbb{X}$ is some locally convex Hausdorff topological space such as $\mathcal{S}(\mathbb{H})$ and $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ or all other spaces related to quantum information considered in this book.

Recall that $\mathcal{A} \subseteq \mathbb{X}$ is said to be convex if for any $x, y \in \mathcal{A}$ then $px+(1-p)y \in \mathcal{A}$ for all $p \in [0,1]$. For any $\mathcal{A} \subseteq \mathbb{X}$, $\text{co}(\mathcal{A})$ denotes the convex hull (i. e., the smallest convex set that contains $\mathcal{A}$) and $\overline{\text{co}}(\mathcal{A})$ is the closed convex hull (i. e., the smallest closed convex set that contains $\mathcal{A}$). For any $x, y, z \in \mathcal{A}$, we say that z lies between x and y if $x \neq y$ and there exists a $0 < p < 1$ such that $z = px + (1-p)y$. $z \in \mathcal{A}$ is said to be an extreme point of $\mathcal{A}$ if it does not lie between any two distinct points of $\mathcal{A}$. That is, if there does not exist $x, y \in \mathcal{A}$ and $0 < p < 1$ such that $x \neq y$ and $z = px + (1-p)y$. The set of all extreme points of $\mathcal{A}$ is denoted by $\text{extr}(\mathcal{A})$.

We have the following theorem regarding a convex and compact set $\mathcal{A} \subseteq \mathbb{X}$. Its proof can be found in Rockafellar [131] and is therefore omitted.

Theorem 3.3.1 (Krein–Millman theorem). *If $\mathcal{A} \subseteq \mathbb{X}$ is convex and compact, then $\mathcal{A}$ has extreme points. Furthermore, $\mathcal{A}$ is the closed convex hull of its extreme points* $\text{extr}(\mathcal{A})$, *i. e.,* $\mathcal{A} = \overline{\text{co}}(\text{extr}(\mathcal{A}))$.

Note that when $\mathcal{A} = \mathcal{S}(\mathbb{H})$ (the space of quantum states on $\mathbb{H}$), $\text{extr}(\mathcal{S}(\mathbb{H}))$ consists of all pure quantum states on the system represented by $\mathbb{H}$, since any pure state cannot be expressed as a nontrivial convex combination of any other quantum states by its definition.

3.3.2 Properties of barycenter

To define barycenter centers of a Borel probability measure on a closed set $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H})$, we need to explore the concept of the Bochner integral. In mathematics, the Bochner integral, named for Salomon Bochner [13] extends the definition of Lebesgue integral to functions that take values in a Banach space, as the limit of integrals of simple functions.

The following definition (Definition 3.3.2) and proposition (Proposition 3.3.3) serve as an introduction to the construction and properties of Bochner integral for readers who are not familiar to them.

Definition 3.3.2. Let $(\mathbb{X}.\Sigma, \mu)$ be a measure space and let $\mathbb{B}$ be a complex Banach space under the Banach norm $\|\cdot\|_{\mathbb{B}}$.

1. A function $f : \mathbb{X} \to \mathbb{B}$ is said to be Σ-measurable if $f^{-1}(B) \in \Sigma$ for all Borel subsets B of $\mathbb{B}$, where

$$f^{-1}(B) = \{x \in \mathbb{X} \mid f(x) \in B\}.$$

2. A Σ-measurable function $f : \mathbb{X} \to \mathbb{B}$ is said to be a simple function if it can be written as

$$f(x) = \sum_{i=1}^{n} 1_{E_i}(x) b_i, \tag{3.8}$$

where $E_i \in \Sigma$ are disjoint sets, b_i are distinct elements in $\mathbb{B}$ and 1_E is the characteristic function of the set $E \in \Sigma$ (i. e., $1_E(x) = 1$ if $x \in E$ and $1_E(x) = 0$ if $x \notin E$).

3. If $\mu(E_i) < +\infty$ for $b_i \neq 0$, then the simple function $f(x) = \sum_{i=1}^{n} 1_{E_i}(x)b_i$, is said to be Bochner integrable and its Bochner integral is defined to be

$$\int_{\mathbb{X}} f(x)\mu(dx) = \int_{\mathbb{X}} \left[\sum_{i=1}^{n} 1_{E_i}(x)b_i \right] \mu(dx) = \sum_{i=1}^{n} b_i \mu(E_i).$$

4. A Σ-measurable function $f : \mathbb{X} \to \mathbb{B}$ is said to be Bochner integrable if there exists a sequence of Bochner integrable simple functions $(f_i)_{i=1}^{+\infty}$ such that

$$\lim_{i \to +\infty} \int_{\mathbb{X}} \|f(x) - f_i(x)\|_{\mathbb{B}} \mu(dx) = 0.$$

In this case, its Bochner integral (with respect to μ) is defined as

$$\int_{\mathbb{X}} f(x)\mu(dx) = \lim_{i \to \infty} \int_{\mathbb{X}} f_i(x)\mu(dx).$$

Many of the familiar properties of the Lebesgue integral continue to hold for the Bochner integral. Particularly useful is Bochner's criterion for integrability, which states that if $(\mathbb{X}, \Sigma, \mu)$ is a measure space, then a Bochner-measurable function $f : \mathbb{X} \to \mathbb{B}$ is Bochner integrable if and only if

$$\int_{\mathbb{X}} \|f(x)\|_{\mathbb{B}} \mu(dx) < +\infty.$$

We state the following proposition without a proof, since we use the concept Bochner integral as a tool only for our exposition of infinite-dimensional quantum information. Interested readers are referred to Bochner [13] and Cohn [27] for the details.

Proposition 3.3.3. *Let* $(\mathbb{X}, \Sigma, \mu)$ *be a measure space and let* $f : \mathbb{X} \to \mathbb{B}$ *be a Bochner integrable function (with respect to* μ*). Then the following properties hold:*

1. *If* $\mathbf{T}$ *is a bounded linear operator on* $\mathbb{B}$*, then* $\mathbf{T}f : \mathbb{B} \to \mathbb{B}$ *is Bochner integrable. Moreover,*

$$\mathbf{T}\left(\int_{\mathbb{X}} f(x)\mu(dx) \right) = \int_{\mathbb{X}} \mathbf{T}f(x)\mu(dx).$$

2. (Dominated convergence theorem). *If* $(f_n)_{n=1}^{+\infty}$, $f_n : \mathbb{X} \to \mathbb{B}$, *is a sequence of* Σ*-measurable function on the complete measure space* $(\mathbb{X}, \Sigma, \mu)$ *converges almost everywhere to a limiting function* $f : \mathbb{X} \to \mathbb{B}$ *and if*

$$\|f(x)\|_{\mathbb{B}} \leq g(x) \text{ for almost every } x \in \mathbb{X},$$

where $g : \mathbb{X} \to \mathbb{R}$ is some Bochner integrable function, then

$$\lim_{n\to+\infty} \int_{\mathbb{X}} \|f(x) - f_n(x)\|_{\mathbb{B}} \mu(dx) = 0$$

and

$$\int_E f(x)\mu(dx) = \lim_{n\to+\infty} \int_E f_n(x)\mu(dx), \quad \forall E \in \Sigma.$$

3. *The following inequality holds:*

$$\left\| \int_E f(x)\mu(dx) \right\|_{\mathbb{B}} \le \int_E \|f(x)\|_{\mathbb{B}} \mu(dx), \quad \forall E \in \Sigma.$$

In particular, the set function

$$E \mapsto \int_E f(x)\mu(dx)$$

define a countably additive $\mathbb{B}$-valued vector measure on $\mathbb{X}$, which is absolutely continuous with respect to μ.

Definition 3.3.4. The barycenter $\bar{\rho}(\mu)$ of $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$ is the Bochner integral defined by

$$\bar{\rho}(\mu) = \int_{\mathcal{S}(\mathbb{H})} \rho\mu(d\rho). \tag{3.9}$$

In particular, if $\mu = \{p_i, \rho_i\}$, then $\bar{\rho}(\mu) = \sum_i p_i\rho_i$ is the average state of the discrete ensemble $\{p_i, \rho_i\}$.

Let $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H})$ be a convex subset and let $\mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}))$ be the collection of Borel probability measures $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$ such that their barycenters $\bar{\rho}(\mu) \in \mathcal{A}$. In particular, if $\mathcal{A} = \{\rho\}$ for some $\rho \in \mathcal{S}(\mathbb{H})$. We write, for notational simplicity,

$$\mathcal{P}_\rho(\mathcal{S}(\mathbb{H})) = \mathcal{P}_{\{\rho\}}(\mathcal{S}(\mathbb{H})) = \{\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H})) \mid \bar{\rho}(\mu) = \rho\}.$$

The following result regarding the barycenter map is due to Holevo and Shirokov [80].

Proposition 3.3.5. *The barycenter map*

$$\mathcal{P}(\mathcal{S}(\mathbb{H})) \ni \mu \mapsto \bar{\rho}(\mu) \in \mathcal{S}(\mathbb{H}) \tag{3.10}$$

is continuous and surjective.

Proof. Let $(\mu_n)_{n=1}^{+\infty} \subset \mathcal{P}(\mathcal{S}(\mathbb{H}))$ be a sequence of Borel probability measures on $\mathcal{S}(\mathbb{H})$ that converges weakly to $\mu_0 \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$. For each $n \in \mathbb{N}$, let $\bar{\rho}(\mu_n) = \int_{\mathcal{S}(\mathbb{H})} \rho\mu_n(d\rho) \in \mathcal{S}(\mathbb{H})$ be the barycenter of μ_n. We consider

$$\begin{aligned}
&\lim_{n\to+\infty} \|\bar{\rho}(\mu_n) - \bar{\rho}(\mu_0)\|_1 \\
&= \lim_{n\to+\infty} \operatorname{tr}\Big[\Big| \int_{\mathcal{S}(\mathbb{H})} \rho\mu_n(d\rho) - \int_{\mathcal{S}(\mathbb{H})} \rho\mu_0(d\rho)\Big|\Big] \\
&= \lim_{n\to+\infty} \int_{\mathcal{S}(\mathbb{H})} \operatorname{tr}[|\rho|]\mu_n(d\rho) - \int_{\mathcal{S}(\mathbb{H})} \operatorname{tr}[|\rho|]\mu_0(d\rho) = 0,
\end{aligned}$$

since $\operatorname{tr}[|\cdot|]$ is a bounded continuous real-valued function on $\mathcal{S}(\mathbb{H})$ and $(\mu_n)_{n=1}^{+\infty}$ converges to μ_0 weakly. This shows that the barycenter map $\bar{\rho}(\cdot) : \mathcal{P}(\mathcal{S}(\mathbb{H})) \to \mathcal{S}(\mathbb{H})$ is a continuous map. It is clear that the barycenter map is an onto map because for each $\rho \in \mathcal{S}(\mathbb{H})$ is a barycenter of a point mass measure at ρ. To show that the barycenter map $\bar{\rho}(\cdot)$ is a one-to-one map by assuming that $\bar{\rho}(\mu) = \bar{\rho}(\nu)$, we want to show that $\mu = \nu$. Note that

$$\bar{\rho}(\mu) = \int_{\mathcal{S}(\mathbb{H})} \rho\mu(d\rho) = \int_{\mathcal{S}(\mathbb{H})} \rho\nu(d\rho) = \bar{\rho}(\nu)$$

implies that $\mu = \nu$ almost everywhere on $\mathcal{S}(\mathbb{H})$. This proves the proposition. □

First, we have the following lemma regarding relationship between the support of $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$ (see Definition 3.1.1 for a definition of support of a measure) and its barycenter $\bar{\rho}(\mu)$.

Lemma 3.3.6. *Let $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$. If U is a closed convex subset of $\mathcal{S}(\mathbb{H})$ such that* $\operatorname{supp}(\mu) \subset U$*, then*

$$\bar{\rho}(\mu) := \int_{\mathcal{S}(\mathbb{H})} \rho\mu(d\rho) \in U. \tag{3.11}$$

Proof. (i) We first assume that $\operatorname{supp}(\mu)$ is a finite set, i. e., there exists a finite set $\{\rho_1, \rho_2, \ldots, \rho_n\} \subset U$ such that $p_i := \mu(\{\rho_i\}) > 0$ and $\sum_{i=1}^n p_i = 1$. In this case, $\bar{\rho}(\mu) = \sum_{i=1}^n p_i\rho_i \in U$, since $\bar{\rho}(\mu)$ is a convex combination of a finite set $\{\rho_1, \rho_2, \ldots, \rho_n\}$ of elements in U.

(ii) We define the set $\mathcal{A}$ by

$$\mathcal{A} = \{\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H})) \mid \operatorname{supp}(\mu) \subset U \Rightarrow \bar{\rho}(\mu) \in U\}. \tag{3.12}$$

We claim that the set $\mathcal{A}$ is dense in $\mathcal{P}(\mathcal{S}(\mathbb{H}))$. Suppose the claim were false for contradiction purposes. Then there exists

$$\mu_0 \in \mathcal{P}(\mathcal{S}(\mathbb{H}))\backslash\mathcal{A} := \mathcal{A}^c$$

and an open neighborhood $\mathcal{O} \subset \mathcal{P}(\mathcal{S}(\mathbb{H}))$ of μ_0 such that $\mathcal{O} \cap \mathcal{A} = \emptyset$. This implies that $\mathrm{supp}(\mu_0)$ is a finite set contained in U, and hence $\bar{\rho}(\mu_0) \in U$ from (i). This is a contradiction.

(iii) We note that $\bar{\rho}(\mathcal{A})$ is dense in $\mathcal{S}(\mathbb{H})$, since it is the image of the continuous map $\mu \mapsto \bar{\rho}(\mu)$ of a dense subset $\mathcal{A}$ of $\mathcal{P}(\mathcal{S}(\mathbb{H}))$. Therefore, $\bar{\rho}(\mathcal{A})$ can be extended to $\bar{\rho}(\mathcal{P}(\mathcal{S}(\mathbb{H})))$. This proves the lemma. □

For $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H})$, we define $\mathrm{co}(\mathcal{A})$, the convex hull of $\mathcal{A}$, the smallest convex set containing $\mathcal{A}$ and $\overline{\mathrm{co}}(\mathcal{A})$, the convex closure of the set $\mathcal{A}$, as the smallest convex closed subset containing $\mathcal{A}$.

Lemma 3.3.7. *Let $\mathcal{A}$ be a closed subset of $\mathcal{S}(\mathbb{H})$. Then $\overline{\mathrm{co}}(\mathcal{A})$ coincides with the set of barycenters of all Borel probability measures supported by $\mathcal{A}$.*

Proof. Let $\rho_0 \in \overline{\mathrm{co}}(\mathcal{A})$. Then there is a sequence $(\rho_n)_{n=1}^{+\infty} \subseteq \mathrm{co}(\mathcal{A})$ converging to ρ_0, so that $(\rho_n)_{n=1}^{+\infty}$ is relatively compact in $\mathcal{S}(\mathbb{H})$. The density operator ρ_n is barycenter of Borel probability measure π_n finitely supported on $\mathcal{A}$. By the compactness criterion of Theorem 3.2.1, the sequence $(\pi_n)_{n=1}^{+\infty}$ is weakly relatively compact, and thus has a partial limit π_0, which is supported by the set $\mathcal{A}$ due to Theorem 3.2.7. Continuity of the map

$$\pi \mapsto \bar{\rho}(\pi) := \int_{\mathcal{S}(\mathbb{H})} \sigma\pi(d\sigma)$$

implies that the state ρ_0 is the barycenter of the measure π_0. Conversely, let π be an arbitrary probability measure supported by $\mathcal{A}$. This measure can be weakly approximated by a sequence of measures $(\pi_n)_{n=1}^{\infty}$ finitely supported by $\mathcal{A}$. Since $\bar{\rho}(\pi_n)$ is in $\mathrm{co}(\mathcal{A})$ for all n, we conclude that $\bar{\rho}(\pi)$ is in $\mathrm{co}(\mathcal{A})$ due to continuity of the map $\pi \mapsto \bar{\rho}(\pi)$. This proves the lemma. □

The following can be found in Lemma 1 of Shirokov [144].

Lemma 3.3.8. *Let $\mathcal{A}$ be a closed subset of $\mathcal{S}(\mathbb{H})$. For any arbitrary state ρ in $\overline{\mathrm{co}}(\mathcal{A})$, there exists a measure μ in $\mathcal{P}(\mathcal{A})$ such that $\bar{\rho}(\mu) = \rho$, where $\mathcal{P}(\mathcal{A})$ is the collection of Borel probability measures on $\mathcal{A}$.*

Proof. Let $\rho \in \overline{\mathrm{co}}(\mathcal{A})$ (where $\mathrm{co}(\mathcal{A})$ denotes the convex hull, i. e., the smallest convex set that contains $\mathcal{A}$, and $\overline{\mathrm{co}}(\mathcal{A})$ is the closed convex hull, i. e., the smallest closed convex set that contains $\mathcal{A}$). Let $(\rho_n)_{n=1}^{+\infty} \subset \mathrm{co}(\mathcal{A})$ be a sequence converging to the state ρ under the $\|\cdot\|_1$-norm. For each $n \in \mathbb{N}$, $\rho_n \in \mathrm{co}(\mathcal{A})$ implies that $\rho_n = \sum_{i=1}^{m(n)} p_{n,i}\rho_{n,i}$ for some $\{p_{n,i}\}_{i=1}^{m(n)}$ and $\{\rho_{n,i}\}_{i=1}^{m(n)} \subset \mathcal{A}$ with $p_{n,i} > 0$ and $\sum_{i=1}^{m(n)} p_{n,i} = 1$, where $m(n)$ is a positive integer that depends on n. That is, $\rho_n = \bar{\rho}(\mu_n)$, where $\mu_n = \{p_{n,i}, \rho_{n,i}\} \in \mathcal{A}$ (i. e., μ_n is a discrete ensemble in $\mathcal{P}(\mathcal{A})$). By the compactness of the set $\mathcal{P}(\mathcal{A})$, the sequence $(\mu_n)_{n=1}^{+\infty}$ has a subsequence $(\mu_{n_k})_{k=1}^{+\infty}$ that converges weakly to $\mu \in \mathcal{P}(\mathcal{A})$. Continuity of the map $\mu \mapsto \bar{\rho}(\mu)$ implies

$$\lim_{k\to+\infty} \rho_{n_k} = \lim_{k\to+\infty} \bar{\rho}(\mu_{n_k}) = \bar{\rho}(\mu) = \rho,$$

where the limit is taken under the $\|\cdot\|_1$. This proves the lemma. □

Let $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$. Recall from Definition 3.1.1 the support of μ denoted by $\operatorname{supp}(\mu)$ is defined by

$$\operatorname{supp}(\mu) = \{\rho \in \mathcal{S}(\mathbb{H}) \mid \mu(N_\rho) > 0 \text{ for every open neighborhood of } \rho\}.$$

Lemma 3.3.9. *The subset of Borel probability measures with finite support is dense in the set of all measures with given barycenter. That is, there exists a* $\rho \in \mathcal{S}(\mathbb{H})$ *such that*

$$\overline{\{\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H})) \mid \sharp(\operatorname{supp}(\mu)) < \infty\}}^{\text{weak}} = \mathcal{P}_{\{\rho\}}(\mathcal{S}(\mathbb{H})),$$

where $\sharp(\operatorname{supp}(\mu))$ *denotes the cardinality (i. e., the number of elements of* $\operatorname{supp}(\mu)$*) and* $\overline{\{\cdots\}}^{\text{weak}}$ *is the closure of* $\{\cdots\}$ *under the weak convergence topology.*

Proof. We follow the proof given in Lemma 1 of Shirokov and Holevo [158] as follows. We first notice from Lemma 3.3.6 that $\operatorname{supp}(\mu) \subset U$, where U is a closed convex subset of $\mathcal{S}(\mathbb{H})$, implies that

$$\bar{\rho}(\mu) := \int_{\mathcal{S}(\mathbb{H})} \rho\mu(d\rho) \in U. \tag{3.13}$$

Now let μ be any arbitrary measure in $\mathcal{P}(\mathcal{S}(\mathbb{H}))$. Since $\mathcal{S}(\mathbb{H})$ is separable, for each $n \in \mathbb{N}$ there exists a sequence $(A_i^n)_{i=1}^{+\infty}$ of Borel subsets of $\mathcal{S}(\mathbb{H})$ of diameters (in trace norm) less than $1/n$ such that $\mathcal{S}(\mathbb{H}) = \bigcup_{i=1}^{+\infty} A_i^n$, $A_i^n \cap A_j^n = \emptyset$ provided that $j \neq i$. Let $m = m(n)$ be a positive integer such that $\sum_{i=m+1}^{+\infty} \mu(A_i^n) < 1/n$. Consider the finite collection of Borel sets $\{\hat{A}_i^n, i = 1, \ldots, m\}$, where $\hat{A}_i^n = A_i^n$ for $i = 1, \ldots, m$ and $\hat{A}_{m+1}^n = \bigcup_{i=m+1}^{+\infty} A_i^n$. We have

$$\bar{\rho}(\mu) := \int_{\mathcal{S}(\mathbb{H})} \rho\mu(d\rho) = \sum_{i=1}^{+\infty} \int_{A_i^n} \rho\mu(d\rho) = \sum_{i=1}^{m+1} \int_{\hat{A}_i^n} \rho\mu(d\rho) = \sum_{i=1}^{m+1} \mu_i^n \rho_i^n, \tag{3.14}$$

where $\mu_i^n := \operatorname{tr}[\int_{\hat{A}_i^n} \rho\mu(d\rho)] = \mu(\hat{A}_i^n)$ and $\rho_i^n = (\mu(\hat{A}_i^n))^{-1} \int_{\hat{A}_i^n} \rho\mu(d\rho)$. Without loss of generality, we can assume that $\mu_i^n > 0$. Let μ^n be the probability measure on $\mathcal{S}(\mathbb{H})$ ascribing the value μ_i^n to the set $\{\rho_i^n\}$. Equality 3.14 implies that $\bar{\rho}(\mu^n) = \bar{\rho}(\mu)$. Since the measure μ^n has finite support for each n, to prove the assertion of the lemma it suffices to show weak convergence of the sequence of measures $(\mu^n)_{n=1}^{+\infty}$ to the measure μ. By the definition of weak convergence of probability measures (see (3.5)), it is sufficient to show that

$$\lim_{n\to+\infty} \int_{\mathcal{S}(\mathbb{H})} f(\rho)\mu^n(d\rho) = \int_{\mathcal{S}(\mathbb{H})} f(\rho)\mu(d\rho)$$

for all bounded uniformly continuous function $f : \mathcal{S}(\mathbb{H}) \to \mathbb{R}$. Let $M_f = \sup_{\rho\in\mathcal{S}(\mathbb{H})}|f(\rho)|$. For arbitrary $\epsilon > 0$, let $n_\epsilon > 0$ be such that $\epsilon n_\epsilon > 2M_f$ and

$$\sup_{\rho\in U(n_\epsilon)} f(\rho) - \inf_{\rho\in U(n_\epsilon)} f(\rho) < \epsilon$$

for the arbitrary closed ball $U(n_\epsilon)$ of diameter $1/n_\epsilon$. By construction, the set $\hat{A}_i^n$ is contained in some ball $U_i(n)$ for each $i = 1, \dots, m$. By (3.13), the state ρ_i^n lies in the same ball $U_i(n)$. Hence, we have

$$\begin{aligned}
&\left| \int_{\mathcal{S}(\mathbb{H})} f(\rho)\mu^n(d\rho) - \int_{\mathcal{S}(\mathbb{H})} f(\rho)\mu(d\rho)\right| \\
&\quad \le \sum_{i=1}^{m+1} \int_{\hat{A}_i^n} |f(\rho) - f(\rho_i)|\mu(d\rho) \\
&\quad \le \epsilon \sum_{i=1}^{m} \mu(\hat{A}_i^n) + 2M_f\mu(\hat{A}_{m+1}^n) < 2\epsilon, \quad \forall n \ge n_\epsilon.
\end{aligned}$$

This proves the lemma. □

Let $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H})$ be a convex subset. That is, if $\rho, \sigma \in \mathcal{A}$ and $\rho \ne \sigma$, then $p\rho + (1-p)\sigma \in \mathcal{A}$ for all $p \in [0,1]$. Recall that $\mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}))$ denotes the collection of Borel probability measures $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$ such that their barycenters $\bar{\rho}(\mu) \in \mathcal{A}$.

We have the following μ-compactness result due originally to Holevo and Shirokov [80].

Theorem 3.3.10. *The set $\mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H})) \subset \mathcal{P}(\mathcal{S}(\mathbb{H}))$ is compact if and only if the set $\mathcal{A} \subset \mathcal{S}(\mathbb{H})$ is compact.*

Proof. ($\Rightarrow$) Let $\mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}))$ be a compact subset of $\mathcal{P}(\mathcal{S}(\mathbb{H}))$. The set $\mathcal{A}$ is the image of $\mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}))$ of the continuous barycenter map $\bar{\rho}(\cdot) : \mathcal{P}(\mathcal{S}(\mathbb{H})) \to \mathcal{S}(\mathbb{H})$ (see Proposition 3.3.5) and is therefore compact.

($\Leftarrow$) Conversely, let $\mathcal{A}$ be a compact subset of $\mathcal{S}(\mathbb{H})$. Then by Theorem 3.2.5, there exists an $\mathfrak{H}$-operator $\mathbf{H}$ and some $h > 0$ such that $\mathrm{tr}[\rho\mathbf{H}] \le h$ for all $\rho \in \mathcal{A}$. For arbitrary $\mu \in \mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}))$, we have

$$\begin{aligned}
\int_{\mathcal{S}(\mathbb{H})} \mathrm{tr}[\rho\mathbf{H}]\mu(d\rho) &= \int_{\mathcal{S}(\mathbb{H})} \lim_{n\to+\infty} \mathrm{tr}[\rho\mathbf{Q}_n\mathbf{H}]\mu(d\rho) \\
&= \lim_{n\to+\infty} \int_{\mathcal{S}(\mathbb{H})} \mathrm{tr}[\rho\mathbf{Q}_n\mathbf{H}]\mu(d\rho) = \lim_{n\to+\infty} \mathrm{tr}\left[\int_{\mathcal{S}(\mathbb{H})} \rho\mathbf{Q}_n\mathbf{H} \right]\mu(d\rho) \\
&= \mathrm{tr}\left[\int_{\mathcal{S}(\mathbb{H})} \rho\mu(d\rho) \lim_{n\to+\infty} (\mathbf{Q}_n\mathbf{H}) \right] = \mathrm{tr}[\bar{\rho}\mathbf{H}] \le h. \qquad (3.15)
\end{aligned}$$

The existence of the integral on the left-hand side and the first equality follows from the monotone convergence theorem (see part 2 of Proposition 3.3.3).

Let $\mathcal{K}_\epsilon = \{\rho \in \mathcal{S}(\mathbb{H}) \mid \mathrm{tr}[\rho\mathbf{H}] \le h\epsilon^{-1}\}$. The set $\mathcal{K}_\epsilon$ is compact for each $\epsilon > 0$ (see Theorem 3.2.5). By (3.15) for any $\mu \in \mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}))$, we have

$$\mu(\mathcal{S}(\mathbb{H})\backslash\mathcal{K}_\epsilon) = \int_{\mathcal{S}(\mathbb{H})\backslash\mathcal{K}_\epsilon} \mu(d\rho) \le \epsilon h^{-1} \int_{\mathcal{S}(\mathbb{H})\backslash\mathcal{K}_\epsilon} \mathrm{tr}[\rho\mathbf{H}]\mu(d\rho) \le \epsilon. \tag{3.16}$$

Therefore, $\mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}))$ is compact by the Prokhorov Theorem 3.2.1. This proves the theorem. □

3.4 Convex functions on $\mathcal{S}(\mathbb{H})$

Throughout the end of this section, $\mathcal{A}$ will be a closed convex subset of $\mathcal{S}(\mathbb{H})$, unless otherwise specified.

3.4.1 σ-convexity and μ-convexity

We first define the convexity and concavity of a function f on $\mathcal{A}$ below.

If the function $f : \mathcal{A} \subseteq \mathcal{S}(\mathbb{H}) \to [-\infty, +\infty]$ does not take the value $-\infty$, then its convexity means that

$$f\Big(\sum_i p_i\rho_i\Big) \le \sum_i p_i f(\rho_i),$$

for all finite discrete ensemble $\{p_i, \rho_i\}$ in $\mathcal{A}$. On the other hand, if $f : \mathcal{A} \subseteq \mathcal{S}(\mathbb{H}) \to [-\infty, +\infty]$ does not take the value $+\infty$ then its concavity means that

$$f\Big(\sum_i p_i\rho_i\Big) \ge \sum_i p_i f(\rho_i),$$

for all finite discrete ensemble $\{p_i, \rho_i\}$ in $\mathcal{A}$.

Therefore, a real-valued function $f : \mathcal{A} \subseteq \mathcal{S}(\mathbb{H}) \to \mathbb{R}$ is concave if $-f$ is convex and vice versa.

A function $f : \mathcal{A} \subseteq \mathcal{S}(\mathbb{H}) \to [-\infty, +\infty]$ is said to be upper bounded (on $\mathcal{A}$) if there exists a real number M such that $f(\rho) \le M$ for all $\rho \in \mathcal{A}$. The function f is said to be lower bounded (on $\mathcal{A}$) if there exists a real number m such that $m \le f(\rho)$ for all $\rho \in \mathcal{A}$. The function f is said to be bounded (on $\mathcal{A}$) if it is both upper and lower bounded (on $\mathcal{A}$). It is called semibounded if it is either upper bounded or lower bounded (on $\mathcal{A}$).

Definition 3.4.1. A semibounded function f on $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H})$ is said to be σ-convex on $\mathcal{A}$ if

$$f\Big(\sum_i p_i\rho_i\Big) \le \sum_i p_i f(\rho_i) \tag{3.17}$$

for all countable (discrete) ensemble $\{p_i, \rho_i\}_i$ of states in $\mathcal{A}$. The function f is said to be σ-concave on $\mathcal{A}$ if $-f$ is σ-convex on $\mathcal{A}$.

Definition 3.4.2. A semibounded Borel measurable function f on $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H})$ is said to be μ-convex on $\mathcal{A}$ if

$$f\Big(\int_{\mathcal{A}} \rho\mu(d\rho)\Big) \le \int_{\mathcal{A}} f(\rho)\mu(d\rho) \tag{3.18}$$

for all probability measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$. The function f is said to be μ-concave on $\mathcal{A}$, if $-f$ is μ-convex on $\mathcal{A}$.

Example 3.3. Let $f : \mathcal{S}(\mathbb{H}) \to [-\infty, +\infty]$ be the lower-bounded Borel measurable function defined by

$$f(\rho) = \begin{cases} 0 & \text{if } \rho \text{ is a state of finite rank;} \\ +\infty & \text{if } \rho \text{ is a state of infinite rank.} \end{cases}$$

Then f is convex but neither σ-convex nor μ-convex.

In the following, we review sufficient conditions for validity of Jensen's inequality (in discrete and integral forms) for convex functions on Banach spaces $\mathbb{X}$ taking values in $[-\infty, +\infty]$. As a simple example showing importance of the conditions in the below propositions one can consider the affine Borel function on the simplex of all probability distributions with countable number of outcomes taking the value 0 on finite rank distributions and the value $+\infty$ on infinite rank distributions. By using Jensen's inequality for finite convex combinations and a simple approximation, it is easy to prove the following assertion.

The following results on Jensen's inequalities regarding convex functions in the setting of $\mathcal{S}(\mathbb{H})$ can be found in Shirokov [143] and [146].

Proposition 3.4.3 (Integral Jensen's inequality). *Let $f : \mathcal{A} \subseteq \mathcal{S}(\mathbb{H}) \to \mathbb{R}$ be a convex function on a closed bounded convex subset $\mathcal{A}$ of $\mathcal{S}(\mathbb{H})$, which is either lower semicontinuous or upper bounded and upper semicontinuous. Then for arbitrary Borel probability measure μ on the set $\mathcal{A}$ the following inequality holds:*

$$f\Big(\int_{\mathcal{A}} \rho\mu(d\rho)\Big) \le \int_{\mathcal{A}} f(\rho)\mu(d\rho) \tag{3.19}$$

Proof. Let μ_0 be an arbitrary probability measure on the set $\mathcal{A}$. Let f be an upper-bounded and upper-semicontinuous function. Then the functional $\mathcal{P}(\mathcal{A}) \ni \mu \mapsto$

$\int_{\mathcal{A}} f(\rho)\mu(d\rho) \in \mathbb{R}$ is upper semicontinuous on the set $\mathcal{P}(\mathcal{A})$ of Borel probability measures on $\mathcal{A}$ endowed with the weak convergence topology (see Billingsley [11]). Let $(\mu_n)_{n=1}^{+\infty}$ be a sequence of measures with finite support and with the same barycenter as the measure μ_0 and weakly converging to the measure μ_0. By convexity of the function f inequality (3.19) holds with $\mu = \mu_n$ for each $n = 1, 2, \ldots$. By upper semicontinuity of the functional $\mathcal{P}(\mathcal{A}) \ni \mu \mapsto \int_{\mathcal{A}} f(\rho)\mu(d\rho) \in \mathbb{R}$ passing to the limit $n \to +\infty$ in this inequality implies inequality (3.19) with $\mu = \mu_0$. Let f be a lower semicontinuous function. By using the arguments from the proof of Lemma 3.3.7, one can show that the function f is either lower bounded or does not take finite values. It is sufficient to consider the first case. Suppose that $\int_{\mathcal{A}} f(\rho)\mu(d\rho) < +\infty$. By applying the construction used in the proof of the same lemma, it is possible to obtain a sequence $(\mu_n)_{n=1}^{+\infty}$ of measures on the set $\mathcal{A}$ with finite support such that

$$\limsup_{n\to+\infty} \int_{\mathcal{A}} f(\rho)\mu_n(d\rho) \le \int_{\mathcal{A}} f(\rho)\mu_0(d\rho)$$

and

$$\limsup_{n\to+\infty} \int_{\mathcal{A}} f(\rho)\mu_n(d\rho) = \int_{\mathcal{A}} f(\rho)\mu_0(d\rho).$$

By convexity of the function f inequality (3.19) holds with $\mu = \mu_n$ for each $n = 1, 2, \ldots$. By lower semicontinuity of the function f passing to the limit $n \to +\infty$ implies inequality (3.19) with $\mu = \mu_0$. This proves the proposition. □

Proposition 3.4.4 (Discrete Jensen's inequality). *Let f be a convex upper-bounded function on a closed convex bounded subset $\mathcal{A}$ of $\mathcal{S}(\mathbb{H})$. Then for arbitrary countable set $\{\rho_i, i = 1, 2, \ldots\} \subset \mathcal{A}$ with the corresponding probability distribution $\{p_i, i = 1, 2, \ldots\}$ the following inequality holds:*

$$f\left(\sum_{i=1}^{+\infty} p_i\rho_i\right) \le \sum_{i=1}^{+\infty} p_i f(\rho_i).$$

Proof. This is the discrete version of Proposition 3.4.3. To prove the proposition, we simply consider the discrete measure μ with atoms at $x_i, i = 1, 2, \ldots$ with $p_i = \mu(\{x_i\}) > 0$ for $i = 1, 2, \ldots$ with $\sum_{i=1}^{+\infty} p_i = 1$. This proves the proposition. □

The following corollary is a consequence of Proposition 3.4.3.

Corollary 3.4.5. *Let f be an affine lower semicontinuous function on a closed bounded convex subset $\mathcal{A}$ of $\mathcal{S}(\mathbb{H})$. Then for arbitrary Borel probability measure μ on the set $\mathcal{A}$ the following equality holds:*

$$f\left(\int_{\mathcal{A}} \rho\mu(d\rho)\right) = \int_{\mathcal{A}} f(\rho)\mu(d\rho). \tag{3.20}$$

3.4.2 Convex hull and convex closure

General theory of convex and variational analysis can be found in the classic monographs by Rockafellar [131] and Rockafella and Wets [132]. Shirokov [143] and [146] has applied these general theory to the setting of convex functions on $\mathcal{S}(\mathbb{H})$. In particular, he developed characterizations of σ-convex hull on these functions.

Recall that a function $f : \mathcal{A} \subseteq \mathcal{S}(\mathbb{H}) \to [-\infty, +\infty]$ is said to be a closed function if $\{(\rho, f(\rho)) \mid \rho \in \mathcal{A}\}$ is a closed subset of $\mathcal{A} \times [-\infty, +\infty]$, where $\mathcal{A}$ is a closed convex subset of $\mathcal{S}(\mathbb{H})$. Consider the subset

$$\operatorname{epi}(f) = \{(\rho, \lambda) \in \mathcal{A} \times [-\infty, +\infty] \mid \lambda \geq f(\rho)\} \tag{3.21}$$

of the set $\mathcal{A} \times [-\infty, +\infty]$. Note that the function f is uniquely determined by the corresponding set $\operatorname{epi}(f)$. Such a function f is called convex if the set $\operatorname{epi}(f)$ is a convex subset of $\mathcal{A} \times [-\infty, +\infty]$ and it is called closed if the set $\operatorname{epi}(f)$ is a closed subset of $\mathcal{A} \times [-\infty, +\infty]$. Each closed function f is lower semicontinuous in the sense that the set defined by the inequality $f(\rho) \leq \lambda$ is a closed subset of $\mathcal{A}$ for arbitrary $\lambda \in [-\infty, +\infty]$, and conversely, each lower semicontinuous function f is closed. It is possible to show that the lower semicontinuity of a function f means that

$$\liminf_{n \to +\infty} f(\rho_n) \geq f(\rho_0)$$

for arbitrary sequence $(\rho_n)_{n=1}^{+\infty}$ converging to ρ_0 under $\|\cdot\|_1$.

The convex hull, $\operatorname{co}(f)$, and the convex closure, $\overline{\operatorname{co}}(f)$, of the function f on a closed convex subset $\mathcal{A}$ of $\mathcal{S}(\mathbb{H})$ are defined as follows.

Definition 3.4.6. For function $f : \mathcal{A} \subseteq \mathcal{S}(\mathbb{H}) \to [-\infty, +\infty]$, define

$$\operatorname{co}(f)(\rho) = \inf_{(\rho,\lambda) \in \operatorname{co}(\operatorname{epi}(f))} \lambda,$$

where the symbol $\operatorname{co}(\operatorname{epi}(f))$ on the right-hand side means the convex hull of (or the smallest convex set containing) the set $\operatorname{epi}(f)$ defined by (3.21). The convex closure $\overline{\operatorname{co}}(f)$ of the function f is defined by

$$\operatorname{epi}(\overline{\operatorname{co}}(f)) = \overline{\operatorname{co}}(\operatorname{epi}(f)),$$

where the symbol $\overline{\operatorname{co}}$ in the right-hand side means the closure of the convex hull of a set. It follows that $\overline{\operatorname{co}}(f)$ is the greatest convex and closed function majorized by f. This implies

$$\overline{\operatorname{co}}(f)(\rho) \leq \operatorname{co}(f)(\rho) \leq f(\rho), \quad \forall \rho \in \mathcal{A} \subseteq \mathcal{S}(\mathbb{H}).$$

Intuitively, we can say that (i) co(f), the convex hull of f on $\mathcal{A}$, as the maximal convex function on $\mathcal{A}$ majorized by f; and (ii) $\overline{\text{co}}(f)$, the convex closure of function f on $\mathcal{A}$, as the maximal closed (lower semicontinuous) convex function on $\mathcal{A}$ majorized by f.

The definition of co(f)(ρ) above is equivalent to having the following representation:

$$\text{co}(f)(\rho) = \inf_{\sum_i p_i\rho_i=\rho} \sum_i p_i f(\rho_i), \text{ where } p_i > 0, \sum_i p_i = 1. \tag{3.22}$$

If f is a continuous function on compact convex set $\mathcal{A} \subset \mathcal{S}(\mathbb{H})$, then $\text{co}(f) = \overline{\text{co}}(f)$.

For arbitrary real valued function f on $\mathfrak{T}_+(\mathbb{H})$, the *Fenchel transform* f^* is a function on the dual space $\mathfrak{T}_+^*(\mathbb{H})(= \mathfrak{B}_+(\mathbb{H}))$ defined by

$$f^*(\mathbf{A}) = \sup_{\rho\in\mathcal{S}(\mathbb{H})} (\text{tr}[\mathbf{A}\rho] - f(\rho)), \quad \mathbf{A} \in \mathfrak{B}_+(\mathbb{H}).$$

By Frenchel theorem (see Rockafellar [131]), it can be easily proved that the function $\overline{\text{co}}(f)$ coincides with the double Frenchet transform of the function f. That is,

$$\overline{\text{co}}(f(\rho)) = f^{**}(\rho) = \sup_{\mathbf{A}\in\mathfrak{B}_+(\mathbb{H})} (\text{tr}[\mathbf{A}\rho] - f^*(\mathbf{A})), \quad \forall\rho \in \mathcal{S}(\mathbb{H}).$$

This implies that in this case $\overline{\text{co}}(f)$ coincides with the upper bound of the set of all affine continuous functions majorized by f. We also note that the convex closure $\overline{\text{co}}(f)(\cdot)$ of a lower-bounded function $f(\cdot)$ on the set $\mathcal{S}(\mathbb{H})$ is defined as the greatest convex lower semicontinuous (closed) function majorized by $f(\cdot)$.

Equivalently, the convex hull co(f) of a semibounded function f on the set $\mathcal{S}(\mathbb{H})$ is defined as the greatest convex function majorized by f (see, e. g., Rockafellar [131]). Therefore,

$$\text{co}(f)(\rho) = \inf_{\{\pi_i,\rho_i\}\in\mathcal{P}^f_{\{\rho\}}} \left(\sum_i \pi_i f(\rho_i)\right), \rho \in \mathcal{S}(\mathbb{H}), \tag{3.23}$$

where the infimum is over all finite discrete ensembles $\{\pi_i, \rho_i\}$ of states with the average state $\sum_i \pi_i\rho_i = \rho$.

The σ-convex hull, denoted by σ-co(f), of a semibounded function f on the set $\mathcal{S}(\mathbb{H})$ is defined as

$$\sigma\text{-co}(f)(\rho) = \inf_{\{\pi_i\rho_i\}\in\mathcal{P}^f_{\{\rho\}}(\mathcal{S}(\mathbb{H}))} \sum_i \pi_i f(\rho_i), \quad \forall\rho \in \mathcal{S}(\mathbb{H}), \tag{3.24}$$

where the infimum is over all countable ensembles $\{\pi_i, \rho_i\}$ of states with the average state ρ.

We have the following immediate observation.

The function $\sigma\text{-co}(f)(\cdot) : \mathcal{S}(\mathbb{H}) \to [-\infty, +\infty]$ is σ-convex. This is because for any countable ensemble $\{\lambda_i, \sigma_i\}$ with the average state σ and any family $\{\{\pi_{ij}, \rho_{ij}\}_j\}_i$ of countable ensembles such that $\sigma_i = \sum_j \pi_{ij}\rho_{ij}$ for all i the countable ensemble $\{\lambda_i\pi_{ij}, \rho_{ij}\}_{ij}$ has the average state σ. Thus, $\sigma\text{-co}(f)(\cdot)$ is the greatest σ-convex function majorized by $f(\cdot)$.

The μ-convex hull of a Borel semibounded function $f(\cdot)$ on the set $\mathcal{S}(\mathbb{H})$, denoted by $\mu\text{-co}(f)(\cdot)$, is defined as

$$\mu\text{-co}(f)(\rho) = \inf_{\mu \in \mathcal{P}_{\{\rho\}}(\mathcal{S}(\mathbb{H}))} \left(\int_{\mathcal{S}(\mathbb{H})} f(\sigma)\mu(d\sigma) \right), \quad \forall \rho \in \mathcal{S}(\mathbb{H}), \tag{3.25}$$

where the infimum is over all probability measures μ with the barycenter ρ.

If the function $\mu\text{-co}(f)(\cdot)$ is Borel measurable and μ-convex, then it is the greatest μ-convex function majorized by $f(\cdot)$.

Stability of the set $\mathcal{S}(\mathbb{H})$ implies the following result.

Proposition 3.4.7 (Shirokov [143, 146]). *Let $f(\cdot)$ be an upper-semicontinuous function on the set $\mathcal{S}(\mathbb{H})$. Then the convex hull* $\text{co}(f)(\cdot)$ *of this function is upper semicontinuous. If, in addition, the function $f(\cdot)$ is upper bounded then*

$$\text{co}(f)(\cdot) = \sigma\text{-co}(f)(\cdot) = \mu\text{-co}(f)(\cdot).$$

Proof. Upper semicontinuity of the function $\text{co}(f)(\cdot)$ can be proved by using the assertion of below, since for an arbitrary sequence $(\rho_n)_{n=1}^{+\infty}$ of states in $\mathcal{S}(\mathbb{H})$, converging to a state ρ_0, this implies existence of such $\mathfrak{H}$-operator $\mathbf{H}$ in the space $\mathbb{H}$ that $\sup_{n\geq 0} \text{tr}[\mathbf{H}\rho_n] < +\infty$. Coincidence of the functions $\text{co}(f)$ and $\mu\text{-co}(f)(\cdot)$ under the condition of upper boundedness of the function f is easily proved by using upper semicontinuity of the functional $\mu \mapsto \int_{\mathcal{S}(\mathbb{H})} f(\rho)\mu(d\rho)$ on the set $\mathcal{P}(\mathcal{S}(\mathbb{H}))$ and density of measures with finite support in the set of all measures with given barycenter. This proves the proposition. □

However, the following example shows that the condition of Proposition 3.4.7 does not imply coincidence of the function $\text{co}(f)(\cdot)$ with the function $\mu\text{-co}(f)(\cdot) = \sigma\text{-co}(f)(\cdot) = \text{co}(f)(\cdot)$.

Example 3.4. Let $f(\cdot) : \mathcal{S}(\mathbb{H}) \to [-\infty, +\infty]$ be the indicator function of a set consisting of one pure state. Then $\mu\text{-co}(f)(\cdot) = f(\cdot)$. Furthermore, $\text{co}(f)(\cdot) = 0$.

4 Completely positive maps

While it is adequate to describe classical physical systems using appropriate positive maps, it is known, however, that these type of maps are not sufficient to describe quantum communication channels. In this chapter, we give a definition of completely positive maps from one C^*-algebra of bounded linear operators to another. Characterizations and relevant properties of completely positive maps in terms of the Stinepring theorem and Kraus representation are also given. It is shown that the positivity condition of maps is weaker than the complete positivity condition in general. However, these two positivity conditions coincide when one of the C^*-algebra is finite-dimensional or when the map under consideration is commutative or *-homomorphism.

4.1 Definitions and properties

Consider the quantum systems A and B represented by separable complex Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively.

Recall that a linear map $\Upsilon : \mathfrak{B}(\mathbb{H}_{\mathbb{A}}) \to \mathfrak{B}(\mathbb{H}_B)$ is said to be positive if $\Upsilon(\mathbf{A}^*\mathbf{A}) \geq \mathbf{O}$ in $\mathfrak{B}(\mathbb{H}_B)$ for every $\mathbf{A} \in \mathfrak{B}(\mathbb{H}_A)$, where $\mathbf{A}^*$ is the adjoint of $\mathbf{A}$. While a map being positive is useful in describing many relevant quantities in commutative physical systems including classical ones, it however fails to do so in quantum physics. Specifically, the map Υ just being positive on $\mathfrak{B}(\mathbb{H}_A)$ is not sufficient in describing the transformation of quantum states from system A to system B, because the open quantum system often interacts with an external quantum system. In the following, we first show the inadequacy of positivity of Υ. Imagine that outside of the system A, there is often another quantum system, that is often referred to environment E, which is represented by Hilbert space $\mathbb{H}_E$ that interacts with $\mathbb{H}_A$. Consider the composite system $\mathbb{H}_{AE} := \mathbb{H}_A \otimes \mathbb{H}_E$ and the extended transformation $\hat{\Upsilon}$, which consists in applying the transformation $\mathbf{T}$ to $\mathbb{H}_A$ and ignoring $\mathbb{H}_E$. That is, applying the transformation Υ to the $\mathbb{H}_A$-part and the identity to $\mathbb{H}_E$-part. In other words, this means considering the mapping

$$\hat{\Upsilon} = \Upsilon \otimes \mathfrak{I}_E : \mathfrak{B}(\mathbb{H}_A \otimes \mathbb{H}_E) \to \mathfrak{B}(\mathbb{H}_A \otimes \mathbb{H}_E),$$

defined by

$$\hat{\Upsilon}(\mathbf{A} \otimes \mathbf{E}) = (\Upsilon \otimes \mathfrak{I}_E)(\mathbf{A} \otimes \mathbf{E}) = \Upsilon(\mathbf{A}) \otimes \mathfrak{I}_E(\mathbf{E}) = \Upsilon(\mathbf{A}) \otimes \mathbf{E},$$

for all $\mathbf{A} \in \mathfrak{B}(\mathbb{H}_A)$ and $\mathbf{E} \in \mathfrak{B}(\mathbb{H}_E)$, where $\mathfrak{I}_E := \mathfrak{I}_{\mathbb{H}_E}$ is the identity operator on $\mathfrak{B}(\mathbb{H}_E)$, i. e., $\mathfrak{I}_E(\mathbf{E}) = \mathbf{E}$ for all $\mathbf{E} \in \mathfrak{B}(\mathbb{H}_E)$. Even though the mapping Υ is positive, but surprisingly enough, the extended mapping $\hat{\Upsilon} = \Upsilon \otimes \mathfrak{I}_E$ does not necessarily preserve positivity as shown in the following example.

https://doi.org/10.1515/9783110788105-004

Example 4.1. Let $\mathbb{H}_A = \mathbb{H}_B = \mathbb{H}$ and let $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ be defined by of linear operator $\mathbf{A}$, i. e.,

$$\Upsilon(\mathbf{A}) = \overline{\mathbf{A}^*}, \quad \forall \mathbf{A} \in \mathfrak{B}(\mathbb{H}),$$

where $\overline{\mathbf{A}^*}$ is the complex conjugate of the adjoint operator of $\mathbf{A}$. We claim that Υ is a positive operator by showing that $\Upsilon(\mathbf{A}^*\mathbf{A}) \geq \mathbf{0}$ for every $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$ as follows:

$$\Upsilon(\mathbf{A}^*\mathbf{A}) = \overline{(\mathbf{A}^*\mathbf{A})^*} = \overline{\mathbf{A}^*\mathbf{A}} = \Upsilon^*(\mathbf{A})\Upsilon(\mathbf{A}) \geq 0.$$

To illustrate that $\hat{\Upsilon}$ is not a positive-preserving map on $\mathfrak{B}(\mathbb{H} \otimes \mathbb{E})$ in a specific example, we let $\mathbb{E} = \mathbb{C}^2$ and consider the algebra $\mathfrak{B}(\mathbb{H} \otimes \mathbb{E})$ as the algebra of 2×2 matrices with coefficients in $\mathfrak{B}(\mathbb{H})$. Then the mapping $\hat{\Upsilon}$ acts as follows:

$$\hat{\Upsilon}\left(\begin{array}{c|c} \mathbf{A}_0^0 & \mathbf{A}_0^1 \\ \hline \mathbf{A}_0^1 & \mathbf{A}_1^1 \end{array}\right) = \left(\begin{array}{c|c} \Upsilon(\mathbf{A}_0^0) & \Upsilon(\mathbf{A}_0^1) \\ \hline \Upsilon(\mathbf{A}_0^1) & \Upsilon(\mathbf{A}_1^1) \end{array}\right).$$

In particular, if $\mathbb{H} = \mathbb{C}^2$, then

$$\hat{\Upsilon}\left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right) = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

is not positivity preserving.

This property is clearly physically unsatisfactory to consider state transformations, which are not state transformations anymore when considered as part of a larger world, even though one does not interact with the environment. Therefore, a concept stronger than that of "positivity" such as "complete positivity" is required in order to adequately describe the phenomena in a quantum world.

To explore the concept of complete positivity, we first assume that $\mathcal{A} \subseteq \mathfrak{L}(\mathbb{H}_A)$ and $\mathcal{B} \subseteq \mathfrak{L}(\mathbb{H}_B)$ are two *-algebras of linear (but not necessary bounded) operators on $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. Recall that a collection of linear operators on a Hilbert space is a *-algebra if it contains the identity operator and is closed under operator addition "+," operator multiplication/convolution "$\circ$" and operator involution "$*$," which is taken to be the adjointness of the operator.

While there are other equivalent definitions (see Chang [24] or Bratteli and Robinson [15]), we give the formal definition of completely positive maps below.

Definition 4.1.1. Let $\mathcal{A}$ and $\mathcal{B}$ be *-algebras of linear operators on $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. A linear map $\Upsilon : \mathcal{A} \to \mathcal{B}$ is said to be n-positive if for any n pairs of operators $(\mathbf{a}_i, \mathbf{b}_i) \in \mathcal{A} \times \mathcal{B}$ $(i = 1, 2, \ldots, n)$ the following inequality holds:

$$\sum_{i=1}^{n} \mathbf{b}_i^* \Upsilon(\mathbf{a}_i^* \mathbf{a}_i)\mathbf{b}_i \geq \mathbf{0}, \tag{4.1}$$

where $\mathbf{a}_i^*$ and $\mathbf{b}_i^*$ are adjoint operators of $\mathbf{a}_i$ and $\mathbf{b}_i$, respectively, and $\mathbf{0}$ is the zero operator in $\mathcal{B}$. The map $\Upsilon : \mathcal{A} \to \mathcal{B}$ is said to be completely positive if it is n-positive for all $n \geq 1$.

Recall from Definition 2.2.5 that a linear map $\pi : \mathcal{A} \to \mathcal{B}$ is a *-homomorphism if, for any $\mathbf{a}, \mathbf{b} \in \mathcal{A}$, $\pi(\mathbf{ab}) = \pi(\mathbf{a})\pi(\mathbf{b})$, $\pi(\mathbf{a}^*) = \pi(\mathbf{a})^*$ and $\pi(\mathbf{I}_A) = \mathbf{I}_B$. The *-homomorphism π is often referred to as a representation of $\mathcal{A}$ into $\mathcal{B}$ (see Definition 2.2.5). The representation π is said to be a *normal* representation if the *-homomorphism is σ-weakly continuous (see Definition 2.1.2 for its definition and Remark 1.6 for the characterizations of σ-weakly continuity). Note that in the case where $\mathcal{A} = \mathcal{B}$ there exists a trivial representation of $\mathcal{A}$ on $\mathcal{B}$ (or simply of $\mathfrak{B}(\mathbb{H})$ when $\mathbb{H}_A = \mathbb{H}_B) = \mathbb{H}$ given by $\pi(\mathbf{A}) = \mathbf{A}$ for all $\mathbf{A} \in \mathcal{A}$. We shall call it the standard representation of $\mathcal{A}$.

The following are some important completely positive linear maps.

Proposition 4.1.2. *Let $\Upsilon : \mathcal{A} \to \mathcal{B}$ be a *-homomorphism. Then Υ is a completely positive and continuous map.*

Proof. 1. For each $n \geq 1$ and every n pairs $(\mathbf{a}_i, \mathbf{b}_i) \in \mathcal{A} \times \mathcal{B}$, $i = 1, 2, \ldots, n$, we need to show that $\sum_{i=1}^{n} \mathbf{b}_i^* \Upsilon(\mathbf{a}_i^* \mathbf{a}_i)\mathbf{b}_i$ is a positive operator on $\mathbb{H}_B$. Through the definition of *-homomorphism and positivity of operators, we have

$$\begin{aligned}\sum_{i=1}^{n} \mathbf{b}_i^* \Upsilon(\mathbf{a}_i^* \mathbf{a}_i)\mathbf{b}_i &= \sum_{i=1}^{n} \mathbf{b}_i^* \Upsilon(\mathbf{a}_i^*)\Upsilon(\mathbf{a}_i)\mathbf{b}_i \\ &= \sum_{i=1}^{n} \mathbf{b}_i^* \Upsilon(\mathbf{a}_i)^*\Upsilon(\mathbf{a}_i)\mathbf{b}_i = \sum_{i=1}^{n} \big(\Upsilon(\mathbf{a}_i)\mathbf{b_i}\big)^* \Upsilon(\mathbf{a}_i)\mathbf{b}_i \geq \mathbf{0}.\end{aligned}$$

Therefore, Υ is n-positive for any $n \geq 1$. This shows that the *-homomorphism $\Upsilon : \mathcal{A} \to \mathcal{B}$ is completely positive.

2. We now want to prove that Υ is a continuous map from $\mathcal{A}$ to $\mathcal{B}$. If $\mathbf{A}$ is a self-adjoint element of $\mathcal{A}$ then, for *-homomorphism $\Upsilon : \mathcal{A} \to \mathcal{B}$, $\Upsilon(\mathbf{A})$ is a self-adjoint element of $\mathcal{B}$. In particular,

$$\|\Upsilon(\mathbf{A})\|_\infty = \sup\{|\lambda| \mid \lambda \in \sigma_p(\Upsilon(\mathbf{A}))\},$$

where $\sigma_p(\Upsilon(\mathbf{A}))$ is the point spectrum (or the set of all eigenvalues) of $\Upsilon(\mathbf{A}) \in \mathcal{B}$. But note that if r belongs to the resolvent set $\rho(\mathbf{A})$ of $\mathbf{A}$ then $\mathbf{A} - r\mathbf{I}_A$ is invertible in $\mathfrak{B}(\mathbb{H}_A)$; hence, $\Upsilon(\mathbf{A} - r\mathbf{I}_A) = \Upsilon(\mathbf{A}) - r\mathbf{I}_B$ is invertible in $\mathfrak{B}(\mathbb{H}_B)$ (see Subsection 1.4 for the definition and properties of resolvent sets). This is to say that r belongs to the resolvent set $\rho(\Upsilon(\mathbf{A}))$. This proves that $\sigma_p(\Upsilon(\mathbf{A})) \subset \sigma_p(\mathbf{A})$ and this gives the estimate

$$\big\|\Upsilon(\mathbf{A})\big\|_\infty \leq \sup\{|\lambda| \mid \lambda \in \sigma_p(\mathbf{A})\} = \|\mathbf{A}\|_\infty.$$

Now if $\mathbf{A}$ is any element of $\mathcal{A} \subset \mathfrak{B}(\mathbb{H}_A)$, then

$$\|\Upsilon(\mathbf{A})\|_\infty^2 = \|\Upsilon^*(\mathbf{A})\Upsilon(\mathbf{A})\|_\infty = \|\Upsilon(\mathbf{A}^*\mathbf{A})\|_\infty \leq \|\mathbf{A}^*\mathbf{A}\|_\infty = \|\mathbf{A}\|_\infty^2.$$

This proves the continuity of the map $\Upsilon : \mathcal{A} \to \mathcal{B}$. Therefore, the proposition follows. □

For the remaining part of this section, it is assumed that $\mathcal{A}$ and $\mathcal{B}$ are two unital C^*-algebras of bounded linear operators on complex separable Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. Recall from Definition 2.2.4 that a C^*-algebra is a *-algebra of bounded linear operators that is complete under the operator norm $\|\cdot\|_\infty$ and satisfies $\|\mathbf{a}^*\mathbf{a}\|_\infty = \|\mathbf{a}\|_\infty^2$ for all $\mathbf{a} \in \mathcal{A}$. The unit operators of $\mathcal{A}$ and $\mathcal{B}$ are to be denoted by $\mathbf{I}_A$ and $\mathbf{I}_B$, respectively.

While the results represented in this chapter are obtained for general C^*-algebras $\mathcal{A} \subseteq \mathfrak{B}(\mathbb{H}_A)$ and $\mathcal{B} \subseteq \mathfrak{B}(\mathbb{H}_B)$, however, for applications to quantum communication, we often assume that $\mathcal{A} = \mathfrak{B}(\mathbb{H}_A)$ and $\mathcal{B} = \mathfrak{B}(\mathbb{H}_B)$.

We give an alternate and yet easier way to verify definition of complete positive maps as follows. For every integer $n \in \mathbb{N}$, let $\mathcal{M}_n$ be the collection of all $n \times n$ complex matrices. Recall an $n \times n$ matrix $A = [a_{ij}]_{i,j=1}^n \in \mathcal{M}_n$ is said to be positive (or positive definite) if the quadratic form $\langle x, Ax\rangle_{\mathbb{C}^n} \geq 0$ for all nonzero vectors $x \in \mathbb{C}^n$, where $\langle x, y\rangle_{\mathbb{C}^N} := x \cdot y$ is the inner product on $\mathbb{C}^n$ defined by

$$\langle x, y\rangle_{\mathbb{C}^n} := x \cdot y = \sum_{i=1}^n \bar{x}_i y_i \quad \forall x = (x_1, x_2, , \dots, x_n), y = (y_1, y_2, \dots, y_n) \in \mathbb{C}^n.$$

Let $\mathcal{A} \subseteq \mathfrak{B}(\mathbb{H}_A)$ be a C^*-algebra. Let $\mathcal{M}_n(\mathcal{A})$ be the collection of $n \times n$ matrices with entries in $\mathcal{A}$.

Explicitly, the space $\mathcal{A} \otimes \mathcal{M}_n$ (similarly for $\mathcal{B} \otimes \mathcal{M}_n$) consists of all $n \times n$ matrices $\mathbf{a} = [\mathbf{a}_{ij}]_{i,j=1}^n$, where $\mathbf{a}_{ij} \in \mathcal{A}$ for all $i, j = 1, 2, \dots, n$. With the obvious matrix multiplication and the *-operation, $\mathcal{A} \otimes \mathcal{M}_n$ is an involutive algebra, i. e.,

$$(\lambda\mathbf{a} + \mu\mathbf{b})_{ij} = \lambda\mathbf{a}_{ij} + \mu\mathbf{b}_{ij}, \quad (\mathbf{ab})_{ij} = \sum_{k=1}^n \mathbf{a}_{ik}\mathbf{b}_{kj} \quad \text{and} \quad (\mathbf{a}^*)_{ij} = \mathbf{a}_{ji}^*, \tag{4.2}$$

for all $\mathbf{a} = [\mathbf{a}_{ij}]_{i,j=1}^n$, $\mathbf{b} = [\mathbf{b}_{ij}]_{i,j=1}^n$ in $\mathcal{A} \otimes \mathcal{M}_n$ and $\lambda, \mu \in \mathbb{C}$.

The following result certifies that we can identify $\mathcal{M}_n(\mathcal{A})$ with the tensor product $\mathcal{A} \otimes \mathcal{M}_n$.

Proposition 4.1.3. *Let $\mathcal{A}$ be a C^*-algebra of bounded linear operators on $\mathbb{H}_A$. Then $\mathcal{M}_n(\mathcal{A})$ is a *-algebra and $\mathcal{M}_n(\mathcal{A})$ and $\mathcal{A} \otimes \mathcal{M}_n$ are *-isomorphic via the mapping*

$$[\mathbf{a}_{ij}]_{i,j=1}^n \mapsto \sum_{i,j=1}^n \mathbf{a}_{ij} \otimes E_{ij},$$

where E_{ij}'s are the matrix in $\mathcal{M}_n$ with ij entry equals to 1 and 0 everywhere else.

Proof. Clearly, the map is linear and it is multiplicative, since for $[\mathbf{a}_{ij}]_{i,j=1}^n$, and $[\mathbf{b}_{ij}]_{i,j=1}^n$ in $\mathcal{M}_n(\mathcal{A})$, we have

$$\begin{aligned}
[\mathbf{a}_{ij}]_{i,j=1}^n[\mathbf{b}_{ij}]_{i,j=1}^n &= \left[\sum_{k=1}^n \mathbf{a}_{ik}\mathbf{b}_{kj}\right]_{i,j=1}^n \mapsto \sum_{i,j=1}^n \left(\sum_{k=1}^n \mathbf{a}_{ik}\mathbf{b}_{kj}\right) \otimes E_{ij} \\
&= \sum_{i,j=1}^n \left(\sum_{k=1}^n \mathbf{a}_{ik}\mathbf{b}_{kj} \otimes E_{ik}E_{kj}\right) = \sum_{i,j,k,s=1}^n (\mathbf{a}_{ik}\mathbf{b}_{sj}) \otimes E_{ik}E_{sj} \\
&= \left(\sum_{i,j=1}^n \mathbf{a}_{ij} \otimes E_{ij}\right)\left(\sum_{i,j=1}^n \mathbf{b}_{ij} \otimes E_{ij}\right).
\end{aligned}$$

We also have

$$([\mathbf{a}_{ij}]_{i,j=1}^n)^* = [\mathbf{a}_{ji}^*]_{i,j=1}^n \mapsto \sum_{i,j=1}^n \mathbf{a}_{ji}^* \otimes E_{ij} = \sum_{i,j=1}^n \mathbf{a}_{ji}^* \otimes E_{ji}^* = \left(\sum_{i,j=1}^n \mathbf{a}_{ij} \otimes E_{ij}\right)^* .$$

This means that the map is a *-homomorphism. Surjectivity of the map follows from the fact that any element of $\mathcal{A}\otimes\mathcal{M}_n$ is of the form $\sum_{i,j=1}^n \mathbf{a}_{ij}\otimes E_{ij}$ for some $\mathbf{a}_{ij} \in \mathcal{A}$. To see the injectivity of the map, let $\sum_{i,j=1}^n \mathbf{a}_{ij} \otimes E_{ij} = \mathbf{0}$. Then

$$(\mathbf{b} \otimes E_{kr})\left(\sum_{i,j=1}^n \mathbf{a}_{ij} \otimes E_{ij}\right)(\mathbf{c} \otimes E_{sm}) = \mathbf{b}\mathbf{a}_{km}\mathbf{c} \otimes E_{km} = \mathbf{0}.$$

Hence, $\mathbf{b}\mathbf{a}_{km}\mathbf{c} = \mathbf{0}$ for all $\mathbf{b}, \mathbf{c} \in \mathcal{A}$ and $1 \le k, m \le n$. Hence, $\mathbf{a}_{km} = \mathbf{0}$ and so $[\mathbf{a}_{ij}]_{i,j=1}^n = \mathbf{0}$ (the zero matrix). This proves the proposition. □

Explicitly the above result (Proposition 4.1.3) states that the space $\mathcal{A} \otimes \mathcal{M}_n$ can be canonically identified with all $n \times n$ matrices $\mathbf{a} = [\mathbf{a}_{ij}]_{i,j=1}^n$, where $\mathbf{a}_{ij} \in \mathcal{A}$ for all $i, j = 1, 2, \ldots, n$.

The algebraic tensor product C^*-algebras $\mathcal{A} \otimes \mathcal{M}_n$ (resp., $\mathcal{B} \otimes \mathcal{M}_n$) can be represented as the C^*-algebras of $n\times n$ complex matrices with entries in $\mathcal{A}$ (resp., $\mathcal{B}$). In fact, every element $\mathbf{x}$ of $\mathcal{A} \otimes \mathcal{M}_n$ can be written in the form

$$\mathbf{x} = \sum_{1\le i,j\le n} \mathbf{x}_{ij} \otimes E_{ij} = \begin{pmatrix} \mathbf{x}_{11} & \mathbf{x}_{12} & \cdots & \mathbf{x}_{1n} \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \cdots & \mathbf{x}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{n1} & \mathbf{x}_{n2} & \cdots & \mathbf{x}_{nn} \end{pmatrix}, \tag{4.3}$$

where E_{ij} denotes the $n \times n$ matrix with all the entries equal 0 except the ijth, which is equal to 1, and $\mathbf{x}_{ij} \in \mathcal{A}$ for all $1 \le i, j \le n$.

Let $\Upsilon : \mathcal{A} \to \mathcal{B}$ be a linear map. For each $n \in \mathbb{N}$, we define the linear map Υ_n that maps $n \times n$ matrices in $\mathcal{A} \otimes \mathcal{M}_n$ to $n \times n$ matrices in $\mathcal{B} \otimes \mathcal{M}_n$ in the following fashion:

$$\begin{pmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} & \dots & \mathbf{a}_{1n} \\ \mathbf{a}_{21} & \mathbf{a}_{22} & \dots & \mathbf{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{n1} & \mathbf{a}_{n2} & \dots & \mathbf{a}_{nn} \end{pmatrix} \mapsto \begin{pmatrix} \Upsilon(\mathbf{a}_{11}) & \Upsilon(\mathbf{a}_{12}) & \dots & \Upsilon(\mathbf{a}_{1n}) \\ \Upsilon(\mathbf{a}_{21}) & \Upsilon(\mathbf{a}_{22}) & \dots & \Upsilon(\mathbf{a}_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \Upsilon(\mathbf{a}_{n1}) & \Upsilon(\mathbf{a}_{n2}) & \dots & \Upsilon(\mathbf{a}_{nn}) \end{pmatrix}. \tag{4.4}$$

In order to give a useful condition equivalent to complete positivity by means of the maps Υ_n, we prove a simple fact on positive elements of $\mathcal{A} \otimes \mathcal{M}_n$ in the case when $\mathcal{A}$ is a C^*-algebra of bounded linear operators on $\mathbb{H}_A$.

Proposition 4.1.4. *Let $\mathcal{A}$ be a C^*-algebra of bounded linear operators on $\mathbb{H}_A$ and let $\mathbf{x} = [\mathbf{x}_{ij}]_{i,j=1}^n$ be an element of $\mathcal{A} \otimes \mathcal{M}_n$. The following conditions are equivalent:*
1. *$\mathbf{x}$ is positive, i. e., $\mathbf{x} = \mathbf{y}^*\mathbf{y}$ for some $\mathbf{y} = [\mathbf{y}_{ij}]_{i,j=1}^n \in \mathcal{A} \otimes \mathcal{M}_n$.*
2. *$\mathbf{x}$ is finite sum of matrices of the form $\sum_{1\le i,j\le n} \mathbf{a}_i^*\mathbf{a}_j \otimes E_{ij}$ with $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathcal{A}$.*
3. *For all $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathcal{A}$, we have $\sum_{1\le i,j\le n} \mathbf{a}_i^*\mathbf{x}_{ij}\mathbf{a}_j \ge \mathbf{0}$.*

Proof. (1) ⇒ (2). Since $\mathbf{x}$ is positive, it can be written in the form $\mathbf{y}^*\mathbf{y}$ with $\mathbf{y} \in \mathcal{A}\otimes\mathcal{M}_n$. Writing $\mathbf{y}$ in the form $\mathbf{y} = \sum_{1\le i,j\le n} \mathbf{y}_{ij} \otimes E_{ij}$, we have

$$\begin{aligned} \mathbf{x} = \mathbf{y}^*\mathbf{y} &= \Big(\sum_{1\le i,j\le n} \mathbf{y}_{ij} \otimes E_{ij}\Big)^* \Big(\sum_{1\le i,j\le n} \mathbf{y}_{ij} \otimes E_{ij}\Big) \\ &= \sum_{k=1}^n \sum_{1\le i,j\le n} \mathbf{y}_{ki}^*\mathbf{y}_{kj} \otimes E_{ij} = \sum_{1\le i,j\le n} \sum_{k=1}^n \mathbf{y}_{ki}^*\mathbf{y}_{kj} \otimes E_{ij} = \sum_{1\le i,j\le n} \mathbf{a}_i^*\mathbf{a}_j \otimes E_{ij}, \end{aligned}$$

where $\mathbf{a}_i^* = \sum_{k=1}^n \mathbf{y}_{ki}^*$ and $\mathbf{a}_j = \sum_{k=1}^n \mathbf{y}_{kj}$. Therefore, $\mathbf{x}$ can be written as finite sum of matrices of the form $\sum_{1\le i,j\le n} \mathbf{a}_i^*\mathbf{a}_j \otimes E_{ij}$ with $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathcal{A}$.

(2) ⇒ (3) is trivial.

(3) ⇒ (1). By representing $\mathcal{A}$ (see Proposition 2.5.2) as a sub-C^*-algebra of $\mathfrak{B}(\mathbb{H}_\mathcal{A})$, (where $\mathbb{H}_\mathcal{A}$ is the Hilbert space that represents $\mathcal{A}$ in GNS representation) and decomposing $\mathbb{H}_\mathcal{A}$ into cyclic orthogonal subspaces (see part 3 of Definition 2.5.3), we may assume that there exists a cyclic vector $u \in \mathbb{H}_\mathcal{A}$ for the representation $(\pi, \mathbb{H}_\mathcal{A})$. That is, $[\pi(\mathcal{A})u] = \mathbb{H}_\mathcal{A}$, where $\pi : \mathcal{A} \to \mathfrak{B}(\mathbb{H}_\mathcal{A})$ is the *-homomorphism defined by $\pi(\mathbf{a}) = \mathbf{a}$ for all $\mathbf{a} \in \mathcal{A}$. Condition (3) then implies the inequality

$$\sum_{1\le i,j\le n} \langle \mathbf{a}_i u, \mathbf{x}_{ij}\mathbf{a}_j u\rangle_{\mathbb{H}_\mathcal{A}} \ge 0.$$

Therefore, since the vector u is cyclic, we have

$$\sum_{1\le i,j\le n} \langle v_i, \mathbf{x}_{ij} v_j\rangle_{\mathbb{H}_\mathcal{A}} \ge 0$$

for all $v_1, \dots, v_n \in \mathbb{H}$. This completes the proof. □

We are now in a position to prove the following proposition.

Proposition 4.1.5. *Let $\mathcal{A}$ and $\mathcal{B}$ be C^*-algebras of bounded linear operators on $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. Let $\Upsilon : \mathcal{A} \to \mathcal{B}$ be a linear map. The following conditions are equivalent:*
1. *Υ is completely positive.*
2. *For every integer $n \geq 1$, the map $\Upsilon_n : \mathcal{A} \otimes \mathcal{M}_n \to \mathcal{B} \otimes \mathcal{M}_n$ defined by (4.4) is positive.*

Proof. The second condition implies the first by parts (2) and (3) of Proposition 4.1.4. Since the operator $\mathbf{x} = \sum_{i,j} \mathbf{a}_i^* \mathbf{a}_j \otimes E_{ij}$ in $\mathcal{A} \otimes \mathcal{M}_n$ is positive, we have that $\Upsilon_n(\mathbf{x}) = \Upsilon_n(\sum_{i,j} \mathbf{a}_i^* \mathbf{a}_j \otimes E_{ij}) = \sum_{i,j} \Upsilon(\mathbf{a}_i^* \mathbf{a}_j) \otimes E_{ij}$ is positive.

Conversely, the first condition implies that $\sum_{i,j} \Upsilon(\mathbf{a}_i^* \mathbf{a}_j) \otimes E_{ij}$ is positive. Therefore, Υ is positive because of conditions (1) and (2) of Proposition 4.1.4. This proves the proposition. □

Proposition 4.1.6. *Let $\Upsilon : \mathcal{A} \to \mathcal{B}$ be a linear map, where $\mathcal{B} = \mathfrak{B}(\mathbb{H}_B)$ is the C^*-algebra of all the bounded linear operators on a Hilbert space $\mathbb{H}_B$. Then Υ is completely positive if and only if for every $n \geq 1$ and every $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{A}$, $u_1, \ldots, u_n \in \mathbb{H}_B$,*

$$\sum_{1 \leq i,j \leq n} \langle u_i, \Upsilon(\mathbf{a}_i^* \mathbf{a}_j) u_j \rangle_{\mathbb{H}_B} \geq 0.$$

Proof. Notice that the C^*-algebra $\mathcal{B} \otimes \mathcal{M}_n$ can be represented as the C^*-algebra of all bounded linear operators on the n-fold direct sum $\mathbb{H}_B \oplus \cdots \oplus \mathbb{H}_B$. Therefore, the above condition is clearly equivalent to positivity of the map Υ_n on $\mathcal{A} \otimes \mathcal{M}_n$ for every $n \geq 1$. This proves the proposition. □

The above proposition can be parallel and rephrased as follows: A linear map $\Upsilon : \mathcal{A} \to \mathcal{B}$ is called completely positive if, for every $n \geq 1$, the map Υ_n on the algebra $\mathcal{A} \otimes \mathcal{M}_n$ of $\mathcal{A}$-valued $n \times n$ matrices to $\mathcal{B}$-valued $n \times n$ matrices defined below are positive, i. e., it maps positive $n \times n$ operator-valued matrices to positive $n \times n$ operator-valued matrices defined by (4.4).

The above results (Proposition 4.1.5 and Proposition 4.1.6) indicate that we can use the following equivalent definition for complete positivity of the linear operator Υ on $\mathfrak{B}(\mathbb{H})$, where $\mathbb{H}_A = \mathbb{H}_B = \mathbb{H}$.

Definition 4.1.7. A linear operator Υ in $\mathfrak{B}(\mathbb{H})$ is completely positive if, for every $n \in \mathbb{N}$, the natural ampliation Υ_n on $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{C}^n$ given by (4.3) is a positive operator.

Recall that a completely positive operator Υ on $\mathfrak{B}(\mathbb{H})$ is said to be a completely positive normal operator if it is σ-weakly continuous.

The following result indicates that a positive linear map $\Upsilon : \mathcal{A} \to \mathcal{B}$ is in fact completely positive when at least one of the C^*-algebras $\mathcal{A}$ and $\mathcal{B}$ is commutative. The result when $\mathcal{B}$ is commutative is due originally to Arveson [2] and the result when $\mathcal{A}$ is commutative is due originally to Stinespring [167]. The proofs are omitted here.

Theorem 4.1.8. *Let $\Upsilon : \mathcal{A} \to \mathcal{B}$ is a positive linear map. If either $\mathcal{A}$ or $\mathcal{B}$ is a commutative C^*-algebra of bounded linear operators, then the positive linear map Υ is completely positive.*

The following simple properties of completely positive maps turn out to be useful. Its proof is trivial and is therefore omitted.

Proposition 4.1.9. *Let $\Upsilon_1 : \mathcal{A} \to \mathcal{B}$ and $\Upsilon_2 : \mathcal{A} \to \mathcal{B}$ be two completely positive maps. Then the map $\Upsilon_1 + \Upsilon_2 : \mathcal{A} \to \mathcal{B}$ is completely positive.*

Proposition 4.1.10. *Let $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$ be C^*-algebras of bounded linear operators and $\Upsilon_1 : \mathcal{A}_1 \to \mathcal{A}_2$ and $\Upsilon_2 : \mathcal{A}_2 \to \mathcal{A}_3$ be two completely positive maps. Then the map $\Upsilon_2 \circ \Upsilon_1 : \mathcal{A}_1 \to \mathcal{A}_3$ is completely positive.*

Proof. It suffices to notice that, for every integer $n \geq 1$, the map $(\Upsilon_2 \circ \Upsilon_1)_n : \mathcal{A}_1 \otimes \mathcal{M}_n \to \mathcal{A}_3 \otimes \mathcal{M}_n$ coincides with the composition $\Upsilon_{2,n} \circ \Upsilon_{1,n}$, where $\Upsilon_{i,n} : \mathcal{A}_i \to \mathcal{A}_i \otimes \mathcal{M}_n$ is the map defined by equation (4.4) for $i = 1, 2$. The positivity of $\Upsilon_{1,n}$ and $\Upsilon_{2,n}$ imply the positivity of $\Upsilon_{2,n} \circ \Upsilon_{1,n} = (\Upsilon_2 \circ \Upsilon_1)_n$ for all $n \geq 1$. Hence, $\Upsilon_2 \circ \Upsilon_1 : \mathcal{A}_1 \to \mathcal{A}_3$ is completely positive by Proposition 4.1.5. □

Proposition 4.1.11. *Let $\mathcal{A}$ be a C^*-algebra of bounded linear operators on $\mathbb{H}_A$ and let $(\Upsilon^{(k)})_{k=1}^{+\infty}$ be a sequence of completely positive linear maps $\Upsilon^{(k)} : \mathcal{A} \to \mathfrak{B}(\mathbb{H}_B)$. Suppose that, for every $\mathbf{a} \in \mathcal{A}$, the sequence $(\Upsilon^{(k)}(\mathbf{a}))_{k=1}^{+\infty}$ converges weakly, i. e.,*

$$\lim_{k\to+\infty} \langle u, \Upsilon^{(k)}(\mathbf{a})v\rangle_{\mathbb{H}_B} \quad \text{exists for all } \mathbf{a} \in \mathcal{A} \text{ and all } u, v \in \mathbb{H}_B.$$

Then the linear map $\Upsilon : \mathcal{A} \to \mathfrak{B}(\mathbb{H}_B)$ defined by

$$\Upsilon(\mathbf{a}) = \lim_{k\to+\infty} \Upsilon^{(k)}(\mathbf{a}), \quad \forall \mathbf{a} \in \mathcal{A},$$

is completely positive.

Proof. By Proposition 4.1.6, it suffices to note that

$$\sum_{1\leq i,j\leq n} \langle u_i, \Upsilon(\mathbf{a}_i^* \mathbf{a}_j)u_j\rangle_{\mathbb{H}_B} = \lim_{k\to+\infty} \sum_{1\leq i,j\leq n} \langle u_i, \Upsilon^{(k)}(\mathbf{a}_i^* \mathbf{a}_j)u_j\rangle_{\mathbb{H}_B} \geq 0,$$

for every integer $n \geq 1$, every $u_1, \dots, u_n \in \mathbb{H}_B$ and every $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathcal{A}$. This shows that the map $\Upsilon : \mathcal{A} \to \mathfrak{B}(\mathbb{H}_B)$ is completely positive. □

4.2 Some technical results

We now present some technical results regarding the implications of various convergence of an infinite series, $\sum_{n=1}^{+\infty} \mathbf{M}_n^* \mathbf{M}_n$, of operators. These results will be used to ensure the good convergence of the series of operators associated with a specific type of mappings that are important in the sections that follow and in Chapter 5.

The proofs of the technical results presented in this section can be found in Attal [3].

– It is recommended that readers skip the proofs of these technical results at the first reading and revisit them when they are needed at a later time. The reader can consult the lecture notes by Attal [3] for proofs omitted.

Lemma 4.2.1. *Let* $\mathbf{M}$ *and* $\mathbf{X}$ *be any bounded linear operators on separable Hilbert space* $\mathbb{H}$, *where* $\mathbf{X}$ *is self-adjoint. Then we have*

$$\mathbf{M}^*\mathbf{X}\mathbf{M} \le \|\mathbf{X}\|_\infty \mathbf{M}^*\mathbf{M}.$$

Proof. Since $\mathbf{X}$ is self-adjoint, $\langle \phi, \mathbf{X}\phi\rangle_{\mathbb{H}}$ is real for all $\phi \in \mathbb{H}$. By the Cauchy–Schwarz inequality (see equation (1.2)), we have

$$\langle \phi, \mathbf{X}\phi\rangle_{\mathbb{H}} \le \|\mathbf{X}\|_\infty \|\phi\|^2_{\mathbb{H}}$$

for all $\phi \in \mathbb{H}$. Let $\phi = \mathbf{M}\psi$, we have from the above inequality

$$\begin{aligned}\langle \psi, \mathbf{M}^*\mathbf{X}\mathbf{M}\psi\rangle_{\mathbb{H}} &= \langle \mathbf{M}\psi, \mathbf{X}\mathbf{M}\psi\rangle_{\mathbb{H}} = \langle \phi, \mathbf{X}\phi\rangle_{\mathbb{H}} \\ &\le \|\mathbf{X}\|_\infty \|\phi\|^2_{\mathbb{H}} = \|\mathbf{X}\|_\infty \|\mathbf{M}\psi\|^2_{\mathbb{H}} = \|\mathbf{X}\|_\infty \langle \psi, \mathbf{M}^*\mathbf{M}\psi\rangle_{\mathbb{H}}.\end{aligned}$$

This implies that

$$\langle \psi, \mathbf{M}^*\mathbf{X}\mathbf{M}\psi\rangle_{\mathbb{H}} \le \|\mathbf{X}\|_\infty \langle \psi, \mathbf{M}^*\mathbf{M}\psi\rangle_{\mathbb{H}}$$

for all $\psi \in \mathbb{H}$. The lemma follows immediately. □

We first describe some conditions below.

– **Condition (weak-M).** The sequence $(\mathbf{M}_n)_{n=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$ is such that the series $\sum_{n=1}^{+\infty} \mathbf{M}_n^*\mathbf{M}_n$ converges weakly to some bounded linear operator $\mathbf{X}$ on $\mathbb{H}$. That is,

$$\lim_{n\to+\infty} \left\langle \phi, \sum_{i=1}^{n} (\mathbf{M}_n^*\mathbf{M}_n - \mathbf{X})\varphi \right\rangle_{\mathbb{H}} = 0, \quad \forall \phi, \varphi \in \mathbb{H}.$$

– **Condition ($\|\mathbf{M}\|_\infty$).** The sequence $(\mathbf{M}_n)_{n=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$ is such that the series $\sum_{n=1}^{+\infty} \mathbf{M}_n^*\mathbf{M}_n$ converges in operator norm $\|\cdot\|_\infty$ to some bounded linear operator $\mathbf{X}$ on $\mathbb{H}$. That is,

$$\lim_{n\to+\infty} \left\| \sum_{i=1}^{n} \mathbf{M}_i^*\mathbf{M}_i - \mathbf{X} \right\|_\infty = 0.$$

It is easy to see that **Condition ($\|\mathbf{M}\|_\infty$)** implies **Condition (weak-M)**.

We now state some technical results that will be used for the remainder of this chapter below.

Proposition 4.2.2. *Let* $(\mathbf{M}_n)_{n=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$ *be a sequence of operators that satisfies* ***Condition (weak* M*)***. *Then the following conclusions hold:*

1. *If* $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$, *then the series* $\sum_{n=1}^{+\infty} \mathbf{M}_n\mathbf{T}\mathbf{M}_n^*$ *is trace-norm convergent and we have*

$$\operatorname{tr}\left[\sum_{n=1}^{+\infty} \mathbf{M}_n\mathbf{T}\mathbf{M}_n^*\right] = \operatorname{tr}[\mathbf{T}\mathbf{X}].$$

2. *If* $\mathbf{Y} \in \mathfrak{B}(\mathbb{H})$, *then the series* $\sum_{n=1}^{+\infty} \mathbf{M}_n^*\mathbf{Y}\mathbf{M}_n$ *converges strongly.*

Furthermore, if the sequence $(\mathbf{M}_n)_{n=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$ *satisfies condition* ($\|\mathbf{M}\|_\infty$), *then the series* $\sum_{n=1}^{+\infty} \mathbf{M}_n^*\mathbf{Z}\mathbf{M}_n$ *is operator-norm convergent for any* $\mathbf{Z} \in \mathfrak{B}(\mathbb{H})$.

Proof. 1. We prove part 1 in the following steps.

(Step 1) As a first step, we assume that $\mathbf{T}$ is positive. We claim that each of the operators $\mathbf{M}_n\mathbf{T}\mathbf{M}_n^*$ is positive and trace class. To prove this claim, we note that since $\mathbf{T}$ is positive, $\mathbf{T} = \mathbf{A}^*\mathbf{A}$ for some operator $\mathbf{A}$. In this case,

$$\mathbf{M}_n\mathbf{T}\mathbf{M}_n^* = \mathbf{M}_n\mathbf{A}^*\mathbf{A}\mathbf{M}_n^* = (\mathbf{A}\mathbf{M}_n^*)^*\mathbf{A}\mathbf{M}_n^* \geq 0.$$

Therefore, $\mathbf{M}_n\mathbf{T}\mathbf{M}_n^*$ is a positive operator. Now $\mathbf{M}_n\mathbf{T}\mathbf{M}_n^*$ is a trace-class operator due to Proposition 1.8.4. This is because $\mathfrak{T}(\mathbb{H})$ is a two-sided ideal in $\mathfrak{B}(\mathbb{H})$ (see Chang [24] or Bratteli and Robinson [15]). That is, $\mathbf{T}\mathbf{A} \in \mathfrak{T}(\mathbb{H})$ and $\mathbf{A}\mathbf{T} \in \mathfrak{T}(\mathbb{H})$ for all $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$ and $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$. Hence, $\mathbf{M}_n, \mathbf{M}_n^* \in \mathfrak{B}(\mathbb{H})$ and $\mathbf{T} \in \mathfrak{T}(\mathbb{H})$ imply that $\mathbf{T}\mathbf{M}_n^*$ and $\mathbf{M}_n\mathbf{T}$ are in $\mathfrak{T}(\mathbb{H})$. The claim that $\mathbf{M}_n\mathbf{T}\mathbf{M}_n^*$ is a trace-class operator follows with a repeated application of Proposition 1.8.4. Also,

$$\operatorname{tr}[\mathbf{M}_n\mathbf{T}\mathbf{M}_n^*] = \operatorname{tr}[\mathbf{T}\mathbf{M}_n\mathbf{M}_n^*] = \operatorname{tr}[\mathbf{T}\mathbf{M}_n^*\mathbf{M}_n].$$

Now put $\mathbf{Y}_n = \sum_{i=1}^{n} \mathbf{M}_i\mathbf{T}\mathbf{M}_i^*$, for all $n \in \mathbb{N}$. For all $n < m$, the operator $\mathbf{Y}_m - \mathbf{Y}_n$ is positive and trace class. Put $\mathbf{X}_n = \sum_{i=1}^{n} \mathbf{M}_i^*\mathbf{M}_i$. Then, for all $n < m$ we have

$$\begin{aligned}\|\mathbf{Y}_m - \mathbf{Y}_n\|_1 &= \operatorname{tr}[|\mathbf{Y}_m - \mathbf{Y}_n|] = \operatorname{tr}[\mathbf{Y}_m - \mathbf{Y}_n] \\ &= \sum_{n<i\leq m} \operatorname{tr}[\mathbf{M}_i\mathbf{T}\mathbf{M}_i^*] = \sum_{n<i\leq m} \operatorname{tr}[\mathbf{T}\mathbf{M}_i^*\mathbf{M}_i] = \operatorname{tr}[\mathbf{T}(\mathbf{X}_m - \mathbf{X}_n)].\end{aligned}$$

Since $\mathbf{T}$ is a positive trace-class operator on $\mathbb{H}$, it can be decomposed as

$$\mathbf{T} = \sum_{i=1}^{+\infty} \lambda_i |e_i\rangle_{\mathbb{H}}\langle e_i|,$$

where $\{\lambda_i\}_{i=1}^{+\infty}$ is the eigenvalues of $\mathbf{T}$ with $\sum_{i=1}^{+\infty}\lambda_i < \infty$, and $\{|e_i\rangle_{\mathbb{H}}\}_{i=1}^{+\infty}$ are the corresponding eigenvectors. In this case,

$$\operatorname{tr}[\mathbf{T}(\mathbf{X}_m - \mathbf{X}_n)] = \sum_{i\in\mathbb{N}} \lambda_i \langle e_i, (\mathbf{X}_m - \mathbf{X}_n)e_i\rangle_{\mathbb{H}}.$$

As the sequence $(\mathbf{X}_n)_{n=1}^{+\infty}$ converge weakly to $\mathbf{X}$, its Cauchy sequence converges weakly. Hence, each of the terms $\langle e_i, (\mathbf{X}_m - \mathbf{X}_n)e_i\rangle_{\mathbb{H}}$ converges to 0 as n and m go to $+\infty$. Each of the sequences $(\mathbf{X}_n e_i)_{n=1}^{+\infty}$ is bounded, since every weakly convergent sequence is bounded. Hence, the terms $\langle e_i, (\mathbf{X}_m - \mathbf{X}_n)e_i\rangle_{\mathbb{H}}$ are all bounded independently of n and m. By Lebesgue's convergence theorem (see, e. g., Rudin [133]) $\|\mathbf{Y}_m - \mathbf{Y}_n\|_1$ tends to 0 when n and m tend to $+\infty$. In other words, the sequence $(\mathbf{Y}_n)_{n=1}^{+\infty}$ is a Cauchy sequence in $\mathfrak{T}(\mathbb{H})$, hence it converges to some trace-class operator $\mathbf{Y}$. Now,

$$\begin{aligned}
\operatorname{tr}[\mathbf{Y}] &= \operatorname{tr}\Big[\lim_{n\to+\infty}\mathbf{Y}_n\Big] = \lim_{n\to+\infty}\operatorname{tr}[\mathbf{Y}_n] = \lim_{n\to+\infty}\operatorname{tr}\Big[\sum_{i=1}^{n}\mathbf{M}_i\mathbf{T}\mathbf{M}_i^*\Big]\\
&= \lim_{n\to+\infty}\sum_{i=1}^{n}\operatorname{tr}[\mathbf{M}_i\mathbf{T}\mathbf{M}_i^*] = \lim_{n\to+\infty}\sum_{i=1}^{n}\operatorname{tr}[\mathbf{T}\mathbf{M}_i\mathbf{M}_i^*]\\
&= \lim_{n\to+\infty}\operatorname{tr}\Big[\mathbf{T}\Big(\sum_{i=1}^{n}\mathbf{M}_i\mathbf{M}_i^*\Big)\Big] = \lim_{n\to+\infty}\operatorname{tr}\Big[\mathbf{T}\Big(\sum_{i=1}^{n}\mathbf{M}_i^*\mathbf{M}_i\Big)\Big]\\
&= \operatorname{tr}\Big[\mathbf{T}\lim_{n\to+\infty}\mathbf{X}_n\Big] = \operatorname{tr}[\mathbf{X}].
\end{aligned}$$

We have thus proved the proposition for positive trace-class operators $\mathbf{T}$.

(Step 2) As a second step, if $\mathbf{T}$ is a self-adjoint trace-class operator, then $\mathbf{T}$ can be decomposed as $\mathbf{T} = \mathbf{T}_+ - \mathbf{T}_-$, where $\mathbf{T}_+$ and $\mathbf{T}_-$ are positive trace-class operators (see Lemma 1.8.3). It is then easy to extend the above result to self-adjoint trace-class operator $\mathbf{T}$. Finally, if $\mathbf{T}$ is any trace-class operator, one can then decompose it as $\mathbf{T} = \mathbf{A} + i\mathbf{B}$, where $\mathbf{A} = (\mathbf{T} + \mathbf{T}^*)/2$, $\mathbf{B} = -\imath(\mathbf{T} - \mathbf{T}^*)/2$ ($\imath = \sqrt{-1}$) are both self-adjoint and trace class. The conclusion of part 1 of the proposition now follows very easily.

2. We first assume that $\mathbf{Y}$ is a positive bounded linear operator on $\mathbb{H}$. Put $\mathbf{X}_n = \sum_{i=1}^{n}\mathbf{M}_i^*\mathbf{M}_i$ and $\mathbf{S}_n(\mathbf{Y}) = \sum_{i=1}^{n}\mathbf{M}_i^*\mathbf{Y}\mathbf{M}_i$. Note that $\mathbf{S}_n(\mathbf{Y})$ is a positive operator, and even more, all of the operators $\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})$ are positive operators, for all $n < m$. Thus, for all $\phi \in \mathbb{H}$, all $n < m$, we have, using functional calculus and Lemma 4.2.1 several times

$$\begin{aligned}
&\|(\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y}))\phi\|_{\mathbb{H}}^2\\
&= \Big\langle \sqrt{\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})}\phi, (\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y}))\sqrt{\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})}\phi\Big\rangle_{\mathbb{H}}\\
&\le \|\mathbf{Y}\|_\infty\Big\langle \sqrt{\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})}\phi, (\mathbf{X}_m - \mathbf{X}_n)\sqrt{\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})}\phi\Big\rangle_{\mathbb{H}}\\
&\le \|\mathbf{Y}\|_\infty\Big\langle \sqrt{\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})}\phi, \mathbf{X}\sqrt{\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})}\phi\Big\rangle_{\mathbb{H}}\\
&\le \|\mathbf{Y}\|_\infty\Big\langle \phi, \sqrt{\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})}\mathbf{X}\sqrt{\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y})}\phi\Big\rangle_{\mathbb{H}}
\end{aligned}$$

$$\le \|\mathbf{Y}\|_\infty \|\mathbf{X}\|_\infty \langle \phi, (\mathbf{S}_m(\mathbf{Y}) - \mathbf{S}_n(\mathbf{Y}))\phi \rangle_{\mathbb{H}}$$
$$\le \|\mathbf{Y}\|_\infty^2 \|\mathbf{X}\|_\infty \langle \phi, (\mathbf{X}_m - \mathbf{X}_n)\phi \rangle_{\mathbb{H}}.$$

By hypothesis, this term converges to 0 as n and m tend to ∞; hence, we have proved the strong convergence of $(\mathbf{S}_n(\mathbf{Y}))$. In the same way as for part 1 of Proposition 4.2.2 we then easily extend this property to bounded self-adjoint operators and then to general bounded operators.

3. The last part of the proposition is proved as follows. Consider the polar decomposition (see Theorem 1.8.11) $\mathbf{Z} = \mathbf{U}|\mathbf{Z}|$ of the bounded linear operator $\mathbf{Z}$, where $|\mathbf{Z}| = \sqrt{\mathbf{Z}^*\mathbf{Z}}$. We then have

$$\left\| \sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{Z} \mathbf{M}_i \right\|_\infty = \left\| \sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}|\mathbf{Z}|\mathbf{M}_i \right\|_\infty . \tag{4.5}$$

But we have, for all $\phi, \psi \in \mathbb{H}$ with $\|\phi\|_{\mathbb{H}} = \|\psi\|_{\mathbb{H}} = 1$,

$$\begin{aligned}
\left| \left\langle \psi, \left(\sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}|\mathbf{Z}|\mathbf{M}_i \right) \phi \right\rangle_{\mathbb{H}} \right| &\le \sum_{i=n}^{m} |\langle \mathbf{U}^* \mathbf{M}_i \psi, |\mathbf{Z}|\mathbf{M}_i \phi \rangle_{\mathbb{H}}| \\
&\le \sum_{i=n}^{m} \|\mathbf{U}^* \mathbf{M}_i \psi\|_{\mathbb{H}} \||\mathbf{Z}|\mathbf{M}_i \phi\|_{\mathbb{H}} \\
&\le \left(\sum_{i=n}^{m} \|\mathbf{U}^* \mathbf{M}_i \psi\|_{\mathbb{H}}^2 \right)^{1/2} \left(\sum_{i=n}^{m} \||\mathbf{Z}|\mathbf{M}_i \phi\|_{\mathbb{H}}^2 \right)^{1/2} \\
&= \left(\sum_{i=n}^{m} \langle \psi, \mathbf{M}_i^* \mathbf{U}\mathbf{U}^* \mathbf{M}_i \psi \rangle_{\mathbb{H}} \right)^{1/2} \left(\sum_{i=n}^{m} \langle \phi, \mathbf{M}_i^* |\mathbf{Z}|^2 \mathbf{M}_i \phi \rangle_{\mathbb{H}} \right)^{1/2} \\
&= \left\langle \psi, \sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}\mathbf{U}^* \mathbf{M}_i \psi \right\rangle_{\mathbb{H}}^{1/2} \left\langle \phi, \sum_{i=n}^{m} \mathbf{M}_i^* |\mathbf{Z}|^2 \mathbf{M}_i \phi \right\rangle_{\mathbb{H}}^{1/2} \\
&\le \left\| \sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}\mathbf{U}^* \mathbf{M}_i \right\|_\infty^{1/2} \left\| \sum_{i=n}^{m} \mathbf{M}_i^* |\mathbf{Z}|^2 \mathbf{M}_i \right\|_\infty^{1/2} .
\end{aligned}$$

This shows the following particular form of the Cauchy–Schwarz inequality in operator norm $\| \cdot \|_\infty$:

$$\left\| \sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}|\mathbf{Z}|\mathbf{M}_i \right\|_\infty \le \left\| \sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}\mathbf{U}^* \mathbf{M}_i \right\|_\infty^{1/2} \left\| \sum_{i=n}^{m} \mathbf{M}_i^* |\mathbf{Z}|^2 \mathbf{M}_i \right\|_\infty^{1/2} .$$

When applied to equation (4.5), this gives

$$\left\|\sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{Z}\mathbf{M}_i\right\|_\infty \le \left\|\sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}\mathbf{U}^* \mathbf{M}_i\right\|_\infty^{1/2} \left\|\sum_{i=n}^{m} \mathbf{M}_i^* |\mathbf{Z}|^2 \mathbf{M}_i\right\|_\infty^{1/2}.$$

But, using Lemma 4.2.1 again, we get

$$\begin{aligned}\left\|\sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}\mathbf{U}^* \mathbf{M}_i\right\|_\infty &= \sup_{\|\phi\|_{\mathbb{H}}=1} \left\langle \phi, \sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{U}\mathbf{U}^* \mathbf{M}_i \phi \right\rangle_{\mathbb{H}} \\ &= \sup_{\|\phi\|_{\mathbb{H}}=1} \sum_{i=n}^{m} \langle \phi, \mathbf{M}_i^* \mathbf{U}\mathbf{U}^* \mathbf{M}_i \phi\rangle_{\mathbb{H}} \le \|\mathbf{U}\mathbf{U}^*\|_\infty \sup_{\|\phi\|_{\mathbb{H}}=1} \sum_{i=n}^{m} \langle \phi, \mathbf{M}_i^* \mathbf{M}_i \phi\rangle_{\mathbb{H}} \\ &= \|\mathbf{U}\mathbf{U}^*\|_\infty \sup_{\|\phi\|_{\mathbb{H}}=1} \left\langle \phi, \sum_{i=n}^{m} (\mathbf{M}_i^* \mathbf{M}_i)\phi \right\rangle_{\mathbb{H}} = \|\mathbf{U}\|_\infty^2 \left\|\sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{M}_i\right\|_\infty.\end{aligned}$$

This gives finally

$$\left\|\sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{Z}\mathbf{M}_i\right\|_\infty \le \|\mathbf{U}\|_\infty \|\mathbf{Z}\|_\infty \left\|\sum_{i=n}^{m} \mathbf{M}_i^* \mathbf{M}_i\right\|_\infty.$$

By hypothesis, the last term tends to 0 when n and m tend to $+\infty$. This means that the sequence $(\sum_{i=0}^{n} \mathbf{M}_i^* \mathbf{Z}\mathbf{M}_i)_{n=1}^{+\infty}$ is a Cauchy; hence, it converges in operator-norm $\|\cdot\|_\infty$. This proves the proposition. □

We state the following propositions without a proof to avoid repetition of the same type of argument by proving each of those series (treated as sequences of partial sums) is a Cauchy sequence in their norms. Their proofs are quite similar to the one above and are left to the readers.

Proposition 4.2.3. *Let $\mathbb{K}$ be an infinite-dimensional separable complex Hilbert space and let $(e_n)_{n=1}^{+\infty}$ be a given orthonormal basis of $\mathbb{K}$, and let $(\mathbf{M}_n)_{n=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$ be a sequence of operators that satisfies* ***Condition (weak(M))****. Then:*

1. *the series $\sum_{n=1}^{+\infty} (\mathbf{M}_n \otimes \mathbf{I}_{\mathbb{K}}) \otimes |e_n\rangle_{\mathbb{K}}$ converges strongly to a bounded linear map $\mathbf{M} : \mathbb{H} \to \mathbb{H} \otimes \mathbb{K}$. Furthermore, the operator $\mathbf{M}$ satisfies the following properties:*

$$\mathbf{M}(h) = \sum_{n=1}^{+\infty} \mathbf{M}_n(h) \otimes |e_n\rangle_{\mathbb{K}}, \quad \forall h \in \mathbb{H}.$$

   In particular, we have $\mathbf{M}_n = \langle e_n|_{\mathbb{K}} \mathbf{M}$. This operator $\mathbf{M}$ satisfies $\mathbf{M}^*\mathbf{M} = \mathbf{X}$ and the series $\mathbf{M}^* = \sum_{n=1}^{+\infty} \langle e_n|_{\mathbb{K}} (\mathbf{M}_n^* \otimes \mathbf{I}_{\mathbb{K}})$, where the series converges weakly.
2. *Conversely, let $\mathbf{M}$ be any bounded linear operator from $\mathbb{H}$ to $\mathbb{H} \otimes \mathbb{K}$. For all $n \in \mathbb{N}$, put $\mathbf{M}_n = \langle e_n|_{\mathbb{K}} \mathbf{M}$. Then the $\mathbf{M}_n$'s are bounded linear operators on $\mathbb{H}$, the sum $\sum_{n=1}^{+\infty} \mathbf{M}_n^* \mathbf{M}_n$ converges strongly to $\mathbf{M}^*\mathbf{M}$ and the operator $\mathbf{M}$ is given by*

$$\mathbf{M} = \sum_{n=1}^{+\infty} (\mathbf{M}_n \otimes \mathbf{I}_{\mathbb{K}})|e_n\rangle_{\mathbb{K}},$$

   where the sum is strongly convergent.

3. *Finally, in any of two cases above, the mapping*

$$\mathbf{T} \mapsto \Lambda(\mathbf{T}) = \sum_{n=1}^{+\infty} \mathbf{M}_n \mathbf{T} \mathbf{M}_n^*,$$

on $\mathfrak{T}(\mathbb{H})$*, is given by* $\Lambda(\mathbf{T}) = \mathrm{tr}_{\mathbb{K}}[\mathbf{M}\mathbf{T}\mathbf{M}^*]$.
In particular, the mapping Λ *is a bounded linear operator on* $\mathfrak{T}(\mathbb{H})$ *with* $\|\Lambda\|_\infty \leq \|\mathbf{M}^*\mathbf{M}\|_\infty$.

Note that part 3 of Proposition 4.2.3 states that the map $\Lambda : \mathfrak{T}(\mathbb{H}) \to \mathfrak{T}(\mathbb{H})$ defined by the series $\Lambda(\mathbf{T}) = \sum_{n=1}^{+\infty} \mathbf{M}_n \mathbf{T} \mathbf{M}_n^*$ can be expressed as $\Lambda(\mathbf{T}) = \mathrm{tr}_{\mathbb{K}}[\mathbf{M}\mathbf{T}\mathbf{M}^*]$.

In the following, we define its associated bounded linear operator Υ on $\mathfrak{B}(\mathbb{H})$ and note that it is the adjoint of Λ.

Proposition 4.2.4. *Let* $(\mathbf{M}_n)_{n=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$ *be a sequence of operators that satisfies* ***Condition (weak(M))****. Let* $\mathbf{Y}$ *be a bounded linear operator on* $\mathbb{H}$ *and consider the linear map* Υ *on* $\mathfrak{B}(\mathbb{H})$ *defined by*

$$\mathbf{Y} \mapsto \Upsilon(\mathbf{Y}) = \sum_{n=1}^{+\infty} \mathbf{M}_n^* \mathbf{Y} \mathbf{M}_n.$$

Then Υ *is a bounded linear operator on* $\mathfrak{B}(\mathbb{H})$*, with* $\|\Upsilon\|_\infty \leq \|\mathbf{X}\|_\infty$*. The operator* Υ *is the dual of the operator* $\Lambda = \mathrm{tr}_{\mathbb{K}}[\mathbf{M}\mathbf{T}\mathbf{M}^*]$ *defined in the preceding proposition. In particular, the operator* Υ *is* **-weakly continuous on* $\mathfrak{B}(\mathbb{H})$ *and* Λ *is the predual map of* Υ*. Finally, using the same operator* $\mathbf{M} : \mathbb{H} \to \mathbb{H}\otimes\mathbb{K}$ *as in Proposition* 4.2.3*, the operator* Υ *is given by*

$$\Upsilon(\mathbf{Y}) = \mathbf{M}^*(\mathbf{Y} \otimes \mathbf{I}_{\mathbb{K}})\mathbf{M}, \quad \forall \mathbf{Y} \in \mathfrak{B}(\mathbb{H}). \tag{4.6}$$

4.3 Stinespring representation

Let $\mathcal{A}$ be a C^*-algebra of bounded linear operators (but not necessarily on any specifically known complex Hilbert space). Recall from GNS representation of $\mathcal{A}$ (see Definition 2.5.1 and Proposition 2.5.2) that if there exists a *-homomorphism $\pi : \mathcal{A} \to \mathbb{K}$ and a vector $\zeta \in \mathbb{K}$ such that $\pi(\mathcal{A})\zeta$ spans $\mathbb{K}$ (i. e., $\mathrm{span}\{\pi(\mathbf{a})\zeta \mid \mathbf{a} \in \mathcal{A}\} = \mathbb{K}$), then the pair $(\pi, \mathbb{K})$ is called a GNS representation of $\mathcal{A}$ on $\mathbb{K}$ and the vector $\zeta \in \mathbb{K}$ is called a cyclic vector.

The following result due to Stinespring [167] gives a representation of completely positive maps. Although the result holds true for any C^*-algebra of bounded linear operators $\mathcal{A}$ on $\mathbb{H}$, however, for representation simplicity of the material, we assume that $\mathcal{A} = \mathfrak{B}(\mathbb{H})$ below.

Theorem 4.3.1 (Stinespring [167]). *A linear map* $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ *is completely positive if and only if it has the form*

$$\Upsilon(\mathbf{a}) = \mathbf{V}^* \pi(\mathbf{a})\mathbf{V}, \quad \forall \mathbf{a} \in \mathcal{A}, \tag{4.7}$$

where $(\pi, \mathbb{K})$ is a GNS representation of $\mathfrak{B}(\mathbb{H})$ on $\mathbb{K}$ for some Hilbert space $\mathbb{K}$, and $\mathbf{V}$ is a bounded linear operator from $\mathbb{H}$ to $\mathbb{K}$. If the map Υ is normal, then the representation π can be chosen to be normal.

Proof. ($\Leftarrow$) We first assume that Υ be a linear map of the form (4.7) and let $[\mathbf{a}_{ij}]_{i,j=1}^n$ be a positive matrix in $\mathfrak{B}(\mathbb{H}) \otimes \mathcal{M}_n$. For all vectors $(u_j)_{j=1}^n$ in $\mathbb{H}$, we have then

$$\begin{aligned} \sum_{1 \le i,j \le n} \langle u_i, \Upsilon(\mathbf{a}_{ij})u_j \rangle_{\mathbb{H}} &= \sum_{1 \le i,j \le n} \langle u_i, \mathbf{V}^* \pi(\mathbf{a}_{ij})\mathbf{V}u_j \rangle_{\mathbb{H}} \\ &= \sum_{1 \le i,j \le n} \langle \mathbf{V}u_i, \pi(\mathbf{a}_{ij})\mathbf{V}u_j \rangle_{\mathbb{H}} \ge 0, \end{aligned}$$

because $\pi : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{K})$ is a *-homomorphism and, therefore, is completely positive by Proposition 4.1.2. This shows that the map $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ is completely positive by Proposition 4.1.6.

($\Rightarrow$) Conversely, suppose that the linear map $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ is completely positive and consider the vector space $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}$, the algebraic tensor product of $\mathfrak{B}(\mathbb{H})$ and $\mathbb{H}$. On space $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}$, we define the bilinear form $(\cdot,\cdot)_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} : (\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}) \times (\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}) \to \mathbb{C}$ by

$$(\mathbf{x}, \mathbf{y})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = \sum_{1 \le i,j \le n} \langle u_i, \Upsilon(\mathbf{a}_i^* \mathbf{b}_j)v_j \rangle_{\mathbb{H}}$$

for $\mathbf{x} = \sum_i \mathbf{a}_i \otimes u_i$ and $\mathbf{y} = \sum_i \mathbf{b}_i \otimes v_i$ in $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}$. Since Υ is completely positive, we have

$$(\mathbf{x}, \mathbf{x})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = \sum_{1 \le i,j \le n} \langle u_i, \Upsilon(\mathbf{a}_i^* \mathbf{a}_j)u_j \rangle_{\mathbb{H}} \ge 0$$

for all $\mathbf{x} = \sum_i \mathbf{a}_i \otimes u_i$ in $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}$. Hence, the bilinear form $(\cdot,\cdot)_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}}$ defined above is positive.

Consider the algebraic homomorphism π_0 defined on $\mathfrak{B}(\mathbb{H})$ with values in the space of linear transformations on $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}$,

$$\pi_0(\mathbf{a})\left(\sum_{i=1}^n \mathbf{a}_i \otimes u_i \right) = \sum_{i=1}^n (\mathbf{a}\mathbf{a}_i) \otimes u_i, \quad \forall \mathbf{a}, \mathbf{a}_i \in \mathfrak{B}(\mathbb{H}), u_i \in \mathbb{H}.$$

It can be easily shown that $\pi_0(\mathbf{ab}) = \pi_0(\mathbf{a})\pi_0(\mathbf{b})$ and $\pi_0(\mathbf{a}^*) = \pi_0(\mathbf{a})^*$ for all $\mathbf{a}, \mathbf{b} \in \mathfrak{B}(\mathbb{H})$. That is, the map $\pi_0 : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{K})$ is a *-homomorphism, where $\mathbb{K}$ is the Hilbert space defined in the following paragraphs.

Notice that, for all $\mathbf{x}, \mathbf{y}$ as above, we have

$$(\mathbf{x}, \pi_0(\mathbf{a})\mathbf{y})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = (\pi_0(\mathbf{a})^* \mathbf{x}, \mathbf{y})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = (\pi_0(\mathbf{a}^*)\mathbf{x}, \mathbf{y})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}}.$$

It follows that, for every $\mathbf{x} \in \mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}$, the linear map

$$\omega : \mathfrak{B}(\mathbb{H}) \to \mathbb{C}, \quad \omega(\mathbf{a}) = (\mathbf{x}, \pi_0(\mathbf{a})\mathbf{x})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}}$$

is a positive linear functional on $\mathfrak{B}(\mathbb{H})$. Therefore, we have

$$\begin{aligned}\|\pi_0(\mathbf{a})\mathbf{x}\|^2_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} &= (\pi_0(\mathbf{a})\mathbf{x}, \pi_0(\mathbf{a})\mathbf{x})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} \\ &= (\mathbf{x}, \pi_0(\mathbf{a})^*\pi_0(\mathbf{a})\mathbf{x})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = (\mathbf{x}, \pi_0(\mathbf{a}^*)\pi_0(\mathbf{a})\mathbf{x})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} \\ &= (\mathbf{x}, \pi_0(\mathbf{a}^*\mathbf{a})\mathbf{x})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = \omega(\mathbf{a}^*\mathbf{a}\mathbf{I}) \le \|\mathbf{a}^*\mathbf{a}\|_\infty \omega(\mathbf{I}) \\ &= \|\mathbf{a}\|^2_\infty \|\mathbf{x}\|^2_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}},\end{aligned} \tag{4.8}$$

where $\|\cdot\|_\infty$ and $\|\cdot\|_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}}$ denote the operator norm in $\mathbb{H}$ and $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}$, respectively. Let $\mathcal{N}$ be the linear subspace of operators $\mathbf{x}$ in $\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}$ such that $(\mathbf{x}, \mathbf{x})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = \|\mathbf{x}\|^2_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = 0$. Since $\mathcal{N}$ is invariant under $\pi_0(\mathbf{a})$ for every $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$ because of (4.8), we can consider the quotient pre-Hilbert space $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}/\mathcal{N}$. Define the prescalar product on $\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}/\mathcal{N}$ by

$$(\mathbf{x} + \mathcal{N}, \mathbf{y} + \mathcal{N})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}} = (\mathbf{x}, \mathbf{y})_{\mathfrak{B}(\mathbb{H})\otimes\mathbb{H}}.$$

Let $\mathbb{K}$ be the Hilbert space obtained by completion of $(\mathfrak{B}(\mathbb{H}) \otimes \mathbb{H})/\mathcal{N}$. By the above construction, the *-homomorphism π_0 extends to a representation π of $\mathfrak{B}(\mathbb{H})$ on $\mathfrak{B}(\mathbb{K})$ such that

$$\pi(\mathbf{a})(\mathbf{x} + \mathcal{N}) = \pi_0(\mathbf{a})\mathbf{x} + \mathcal{N}$$

for $\mathbf{a} \in \mathfrak{B}(\mathbb{H})$ and $\mathbf{x} \in \mathfrak{B}(\mathbb{H}) \otimes \mathbb{H}$. Consider the linear operator $\mathbf{V} : \mathbb{H} \to \mathbb{K}$,

$$\mathbf{V}u = \mathbf{I} \otimes u + \mathcal{N}.$$

This operator is bounded because of the inequality

$$\|\mathbf{V}u\|^2_{\mathbb{K}} = \langle u, \Upsilon(\mathbf{I})u\rangle_{\mathbb{H}} \le \|\Upsilon(\mathbf{I})\|_\infty \|u\|^2_{\mathbb{H}}.$$

A straightforward computation yields (4.7). This proves the theorem. □

A minimal Stinespring representation (see Chang [24] or Bartteli and Robinson [15]) is defined below.

Definition 4.3.2. A triple $(\mathbb{K}, \pi, \mathbf{V})$ satisfying (4.7) is called a Stinespring representation of the completely positive map $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$. It is called a minimal Stinespring representation if

$$\mathbb{K} = \overline{\{\pi(\mathbf{a})\mathbf{V}u \mid \mathbf{a} \in \mathcal{A}, u \in \mathbb{H}\}}^{\|\cdot\|_{\mathbb{K}}}, \tag{4.9}$$

where $\overline{\{\cdots\}}^{\|\cdot\|_{\mathbb{K}}}$ denotes the closure of $\{\cdots\}$ under $\|\cdot\|_{\mathbb{K}}$.

Completely positive maps describe the dynamics of open quantum systems and Stinespring's dilation theorem is the basic structure theorem for such maps. It states that any completely positive linear map $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ between two C^*-algebras can be written as a concatenation of two basic completely positive linear maps: a *-homomorphism $\pi : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{K})$ into a larger (dilated) C^*-algebra $\mathfrak{B}(\mathbb{K})$ of the bounded operators on some Hilbert space $\mathbb{K}$, followed by a compression $\mathbf{V}^*(\cdot)\mathbf{V}$ into the range algebra $\mathfrak{B}(\mathbb{H})$:

$$\Upsilon(\mathbf{a}) = \mathbf{V}^*\pi(\mathbf{a})\mathbf{V}, \quad \forall \mathbf{a} \in \mathfrak{B}(\mathbb{H}).$$

The Stinespring's theorem provides an excellent characterization of the set of permissible quantum operations (see Definition 5.1.2 for a definition of quantum operation) and is also among the most useful tools in the theory of open quantum systems and quantum information. In a way, the increased system size is the price one has to pay for a simpler description of the map Υ in terms of basic operations $\mathbf{V}$, π and $\mathbf{V}^*$.

Note that every completely positive linear map admits a minimal Stinespring representation. In fact, with the notation of the proof of Theorem 4.3.1, it suffices to consider as Hilbert space $\mathbb{K}$ the closure $\mathbb{K}_1$ of the vector space generated by Theorem 4.3.1. The restriction π_1 of π to $\mathbb{K}_1$ also satisfies (4.9).

The minimal Stinespring representation is unique in the following sense.

Proposition 4.3.3. *Let π_1 and π_2 be two representations of $\mathfrak{B}(\mathbb{H})$ on Hilbert spaces $\mathbb{K}_1$ and $\mathbb{K}_2$ and let $\mathbf{V}_i : \mathbb{H} \to \mathbb{K}_i$, $(i = 1, 2)$, be two bounded linear operators such that*

$$\{\pi_i(\mathbf{a})\mathbf{V}_i u \mid \mathbf{a} \in \mathfrak{B}(\mathbb{H}), u \in \mathbb{H}\},$$

is the total in $\mathbb{K}_i$ for $i = 1, 2$, i. e.,

$$\overline{\{\pi_i(\mathbf{a})\mathbf{V}_i u \mid \mathbf{a} \in \mathfrak{B}(\mathbb{H}), u \in \mathbb{H}\}}^{\|\cdot\|_{\mathbb{K}_i}} = \mathbb{K}_i,$$

and such that

$$\Upsilon(\mathbf{a}) = \mathbf{V}_i^*\pi(\mathbf{a})\mathbf{V}_i$$

for $i = 1, 2$. Then there exists a unitary map $\mathbf{U} : \mathbb{K}_1 \to \mathbb{K}_2$ such that

$$\mathbf{U}\mathbf{V}_1 = \mathbf{V}_2, \quad \mathbf{U}\pi_1(\mathbf{a}) = \pi_2(\mathbf{a})\mathbf{U}, \quad \forall \mathbf{a} \in \mathfrak{B}(\mathbb{H}). \tag{4.10}$$

Proof. Let $\mathbf{U} : \mathbb{K}_1 \to \mathbb{K}_2$ be the densely defined linear map defined by

$$\mathbf{U}\left(\sum_{j=1}^{n} \pi_1(\mathbf{a}_j)\mathbf{V}_1 u_j\right) = \sum_{j=1}^{n} \pi_2(\mathbf{a})\mathbf{V}_2 u_j$$

for every integer $n \geq 1$ and $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathfrak{B}(\mathbb{H})$, $u_1, \ldots, u_n \in \mathbb{H}$. A straightforward computation shows that

$$\begin{aligned}
&\langle \mathbf{U}\pi_1(\mathbf{b})\mathbf{V}_1 v, \mathbf{U}\pi_1(\mathbf{a}_j)\mathbf{V}_1 u\rangle_2 \\
&\quad = \langle v, \Upsilon(\mathbf{b}^*\mathbf{a})u\rangle_{\mathbb{H}} \\
&\quad = (\mathbf{V}_1 v, \pi_1(\mathbf{b}^*\mathbf{a})\mathbf{V}_1 u)_1 \\
&\quad = \langle \pi_1(\mathbf{b})\mathbf{V}_1 v, \pi_1(\mathbf{a})\mathbf{V}_1 u\rangle_1, \quad \forall \mathbf{a}, \mathbf{b} \in \mathfrak{B}(\mathbb{H}) \text{ and } u, v \in \mathbb{H},
\end{aligned}$$

where $(\cdot,\cdot)_j$ denotes the Hilbertian inner product in $\mathbb{K}_i$ for $i = 1, 2$, and naturally $\langle \cdot,\cdot\rangle_{\mathbb{H}}$ the Hilbertian inner product in $\mathbb{H}$. Therefore, $\mathbf{U}$ is an isometry and can be extended to $\mathbb{K}_1$ by an obvious density argument. In a similar way, one can prove that also $\mathbf{U}^* : \mathbb{K}_2 \to \mathbb{K}_1$ is an isometry. Thus, $\mathbf{U}$ is unitary. Finally, since

$$\mathbf{U}\mathbf{V}_1 u = \mathbf{U}\pi_1(\mathbf{1})\mathbf{V}_1 u = \pi_2(\mathbf{1})\mathbf{V}_2 u = \mathbf{V}_2 u, \quad \mathbf{U}\pi_1(\mathbf{a})\mathbf{V}_1 u = \pi_2(\mathbf{a})\mathbf{V}_2 u$$

for every $u \in \mathbb{H}$, (4.10) follows. □

The Stinespring representation of a completely positive map $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ can be considered to be a dilation of Υ. What is a dilation? To explain this concept, let $\mathbb{K}$ be a Hilbert space and $\mathbb{H}$ be a Hilbert subspace of $\mathbb{K}$. If $\mathbf{U}$ is in $\mathfrak{B}(\mathbb{K})$, then $\mathbf{P}_{\mathbb{H}}\mathbf{U}|_{\mathbb{H}} \in \mathfrak{B}(\mathbb{H})$, where $\mathbf{P}_{\mathbb{H}}$ is the projection onto $\mathbb{H}$ and $\mathbf{U}|_{\mathbb{H}}$ is the restriction of $\mathbf{U}$ to $\mathbb{H}$. Set $\mathbf{T} = \mathbf{P}_{\mathbb{H}}\mathbf{U}|_{\mathbb{H}}$. Then $\mathbf{U}$ is said to be a dilation of $\mathbf{T}$ and $\mathbf{T}$ is said to be compression of $\mathbf{U}$. Certainly, any $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$ has many dilations in $\mathfrak{B}(\mathbb{K})$. For example, it can be shown that a contraction has an isometric dilation and a isometry has a unitary dilation.

We summarize Theorem 4.3.1 and Proposition 4.3.3 and obtain the following Stinespring theorem.

Theorem 4.3.4 (Stinespring dilation theorem). *Let $\mathbb{H}$ be a separable complex Hilbert space. If $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ is a completely positive operator, then there exists a Stinespring triple $(\mathbb{K}, \pi, \mathbf{V})$, where $\mathbb{K}$ is a complex Hilbert space, $\mathbf{V} : \mathbb{H} \to \mathbb{K}$ a bounded linear operator with $\|\mathbf{V}\|_\infty^2 = \|\Upsilon(\mathbf{I})\|_\infty$, and $\pi : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{K})$ a *-homomorphism satisfying $\mathbf{V}^*\pi(\cdot)\mathbf{V} = \Upsilon(\cdot)$ on $\mathfrak{B}(\mathbb{H})$. If triple $(\mathbb{K}, \pi, \mathbf{V})$ is minimal, i. e., $\mathbb{K} = \overline{\{\pi(\mathbf{a})\mathbf{V}u \mid \mathbf{a} \in \mathfrak{B}(\mathbb{H}), u \in \mathbb{H}\}}^{\|\cdot\|_{\mathbb{K}}}$, then the triple $(\mathbb{K}, \pi, \mathbf{V})$ becomes unique up to unitary equivalence.*

Remark 4.1. If $\pi : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{K})$ is a unital *-homomorphism and $\mathbf{V} \in \mathfrak{B}(\mathbb{H}, \mathbb{K})$, then the map $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ defined by $\Phi(\mathbf{a}) = \mathbf{V}^*\pi(\mathbf{a})\mathbf{V}$ is completely positive. So, the Stinespring's dilation theorem characterizes the completely positive maps.

Remark 4.2. Let $\mathbb{H}_A$ and $\mathbb{H}_B$ be two complex Hilbert spaces. If $\Upsilon : \mathfrak{B}(\mathbb{H}_A) \to \mathfrak{B}(\mathbb{H}_B)$ is a completely positive map, and if Υ is unital, we may assume that $\mathbb{H}_B$ contains $\mathbb{H}_A$ as a sub-Hilbert space. Indeed, the identity operator $\mathbf{I}_B$ on $\mathbb{H}_B$ can be written as

$$\mathbf{I}_B := \mathbf{I}_{\mathbb{H}_B} = \Upsilon(\mathbf{I}_A) = \mathbf{V}^*\pi(\mathbf{I}_A)\mathbf{V} = \mathbf{V}^*\mathbf{V},$$

where $\mathbf{I}_A := \mathbf{I}_{\mathbb{H}_A}$ (the identity operator on $\mathbb{H}_A$). This implies that $\mathbf{V}$ is an isometry. So, instead of $\mathbb{H}_B = \mathbf{V}(\mathbb{H}_A) \oplus \mathbf{V}(\mathbb{H}_A)^\perp$ we may consider $\mathbb{H}_B = \mathbb{H}_A \oplus \mathbf{V}(\mathbb{H}_A)$. Thus, we have $\Upsilon(\mathbf{a}) = \mathbf{P}_{\mathbb{H}_A}\pi(\mathbf{a})|_{\mathbb{H}_A}$ for all $\mathbf{a} \in \mathfrak{B}(\mathbb{H}_A)$. That is, any completely positive unital map is a compression of a unital *-homomorphism.

Remark 4.3. When the C^*-algebra $\mathcal{A}$ and the Hilbert space $\mathbb{H}$ are separable, then we may assume that $\mathbb{K}$ is separable. Similarly, when $\mathcal{A}$ and $\mathbb{H}$ are finite-dimensional then $\mathbb{K}$ may be taken finite-dimensional.

Remark 4.4. Given a Stinespring representation $(\mathbb{K}, \pi, \mathbf{K})$ for $\Upsilon : \mathcal{A} \to \mathfrak{B}(\mathbb{H})$, it is possible to make it minimal. Let $\mathbb{K}_1$ be the closed linear span of $\pi(\mathbf{A})\mathbf{V}(\mathbb{H})$ in $\mathbb{K}$. Since π is unital, $\mathbf{V}(\mathbb{H})$ lies in $\mathbb{K}_1$ so we may assume that $\mathbf{V} : \mathbb{H} \to \mathbb{K}_1$. Also, $\pi(\mathbf{a})(\mathbb{K}_1)$ lies in $\mathbb{K}_1$ for all $\mathbf{a} \in \mathcal{A}$ since π is closed under multiplicative and continuous. So $\pi_1 : \mathcal{A} \to \mathfrak{B}(\mathbb{K}_1)$ defined by $\pi_1(\mathbf{a}) = \pi(\mathbf{a})|_{\mathbb{K}_1}$ is well-defined and still a unital *-homomorphism. It is easy to see that $(\mathbb{K}_1, \pi_1, \mathbf{V})$ is a minimal Stinespring representation for Υ.

4.4 Kraus representation

In this section, we explore and prove the Kraus version of the Stinespring theorem (see Kraus [98]). Kraus representation provides a characterization of σ-weakly continuous (i. e., normal) completely positive maps.

The presentation of Kraus representation in this section is largely based on the lecture note by Attal [3].

Recall that a map Υ on a von Neumann algebra $\mathcal{A}$ is said to be normal if $\Upsilon(\vee_\alpha \mathbf{a}_\alpha) = \vee_\alpha \Upsilon(\mathbf{a}_\alpha)$ for all increasing net $(\mathbf{a}_\alpha)_{\alpha \in L}$ that is bounded above where $\vee_\alpha(\cdot)$ denotes the least upper bound.

We first need the following preliminary result.

Lemma 4.4.1. *Let π be a normal representation of $\mathfrak{B}(\mathbb{H})$ on a separable Hilbert space $\mathbb{K}$. Then there exists a direct sum decomposition of $\mathbb{K}$,*

$$\mathbb{K} = \bigoplus_{n=1}^{+\infty} \mathbb{K}_n,$$

where the subspaces $\mathbb{K}_n$ are invariant under π (i. e., $\pi(\mathbb{K}_n) \subset \mathbb{K}_n$) and $\pi|_{\mathbb{K}_n}$, the restriction of π to each $\mathbb{K}_n$, is unitarily equivalent to the standard representation of $\mathfrak{B}(\mathbb{H})$.

Proof. If ϕ is a unit vector in $\mathbb{H}$, then the projection $\mathbf{P} = \pi(|\phi\rangle_{\mathbb{H}}\langle\phi|)$ is nonzero. This is because if $\mathbf{U}_n$ are unitary operators in $\mathfrak{B}(\mathbb{H})$ such that $\phi_n = \mathbf{U}_n\phi$ form a maximal orthonormal set in $\mathbb{H}$ and if $\mathbf{P} = \mathbf{0}$ then

$$\pi(|\phi_n\rangle_{\mathbb{H}}\langle\phi_n|) = \pi(\mathbf{U}_n)\mathbf{P}\pi(\mathbf{U}_n) = \mathbf{0}$$

and, by normality of π,

$$\pi(\mathbf{I}) = \sum_n \pi\big(|\phi_n\rangle_{\mathbb{H}}\langle\phi_n|\big) = \mathbf{0}$$

instead of being equal to $\mathbf{I}$ as it should be. Therefore, $\mathbf{P} \neq \mathbf{0}$. If ψ is a unit vector in $\mathbb{K}$ such that $\mathbf{P}\psi = \psi$, then

$$\begin{aligned}\langle\psi, \pi(\mathbf{X})\psi\rangle_{\mathbb{K}} &= \langle\mathbf{P}\psi, \pi(\mathbf{X})\mathbf{P}\psi\rangle_{\mathbb{K}} = \langle\psi, \pi\big(|\phi\rangle_{\mathbb{H}}\langle\phi|\mathbf{X}|\phi\rangle_{\mathbb{H}}\langle\phi|\big)\psi\rangle_{\mathbb{K}} \\ &= \langle\phi, \mathbf{X}\phi\rangle_{\mathbb{H}}, \langle\psi, \pi\big(|\phi\rangle_{\mathbb{H}}\langle\phi|\big)\psi\rangle_{\mathbb{K}} = \langle\phi, \mathbf{X}\phi\rangle_{\mathbb{H}}.\end{aligned}$$

Consider the sub-Hilbert space of $\mathbb{K}$,

$$\mathbb{K}_1 = \overline{\{\pi(\mathbf{X})\phi \mid \mathbf{X} \in \mathfrak{B}(\mathbb{H})\}}.$$

It is clear that $\mathbb{K}_1$ is stable (i. e., closed) under all the operators $\pi(\mathbf{A})$. Hence, so is the space $\mathbb{K}_1^{\perp}$. Consider the map $\mathbf{U} : \pi(\mathbf{X})\psi \to \mathbf{X}\phi$, it is easy to see that $\mathbf{U}$ is isometric, and hence extends into a unitary operator from $\mathbb{K}_1$ to $\mathbb{H}$. Computing $\mathbf{U}\pi(\mathbf{Y})\mathbf{U}^*$ on elements of the form $\mathbf{X}\phi$, shows that $\mathbf{U}\pi(\mathbf{Y})\mathbf{U}^* = \mathbf{Y}$. We have decomposed $\mathbb{K}$ into $\mathbb{K}_1 \oplus \mathbb{K}_1^{\perp}$, where $\mathbb{K}_1$ is stable under π and on which π is unitarily equivalent to the standard representation of $\mathbb{H}$. We are left with a normal representation on $\mathbb{K}_1^{\perp}$. We can apply the same procedure repeatedly and generate a sequence of subspaces $(\mathbb{K}_n)_{n=1}^{+\infty}$ of subspaces of $\mathbb{K}$. By Zorn's lemma (see Lemma 2.1.3), there exists a maximal class of subspaces $\mathbb{K}_n$ of $\mathbb{K}$ with unitary maps $\mathbf{U}_n : \mathbb{K}_n \to \mathbb{H}$ such that $\mathbb{K}_n$ is invariant by π and $\mathbf{X} = \mathbf{U}_n\pi(\mathbf{X})\mathbf{U}_n^*$ for all $\mathbf{X} \in \mathfrak{B}(\mathbb{H})$. This maximal family must satisfy

$$\mathbb{K} = \bigoplus_{n=1}^{+\infty} \mathbb{K}_n$$

for otherwise we can repeat the same construction as for $\mathbb{K}_1$ inside the space

$$\left(\bigoplus_{n=1}^{+\infty} \mathbb{K}_n\right)^{\perp}$$

and contradict the maximality. This proves the lemma. □

The following Kadison–Schwarz inequality (also known as multiplicative inequality) is a generalization of Cauchy–Schwartz inequality (1.2) to the completely positive linear maps on C^*-algebra.

Lemma 4.4.2 (Kadison–Schwarz inequality [96]). *$\mathcal{A}$, $\mathcal{B}$ are unital C^*-algebras and $\Upsilon : \mathcal{A} \to \mathcal{B}$ is a completely positive linear map with $\Upsilon(\mathbf{I}) = \mathbf{I}$, then we have the following Schwarz inequality:*

$$\Upsilon(\mathbf{a}^*)\Upsilon(\mathbf{a}) \leq \Upsilon(\mathbf{a}^*\mathbf{a}), \quad \forall \mathbf{a} \in \mathcal{A}. \tag{4.11}$$

Proof. Let Λ be defined as

$$\Lambda = \mathbf{I} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \mathbf{a} \otimes \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{a} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Then $\Lambda^* \Lambda \geq \mathbf{0}$ and

$$\Lambda^* \Lambda = \begin{bmatrix} \mathbf{I} & \mathbf{a} \\ \mathbf{a}^* & \mathbf{a}^* \mathbf{a} \end{bmatrix}.$$

Now

$$(\Upsilon \otimes \mathbf{I}_2)(\Lambda^* \Lambda) = \begin{bmatrix} \Upsilon(\mathbf{I}) & \Upsilon(\mathbf{a}) \\ \Upsilon(\mathbf{a}^*) & \Upsilon(\mathbf{a}^* \mathbf{a}) \end{bmatrix} \geq \mathbf{0},$$

where

$$\mathbf{I}_2 = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

Now for any vector $\mathbf{U}$ in the tensor space we have that $\langle \mathbf{U}^* (\Upsilon \otimes \mathbf{I}_2), (\Lambda^* \Lambda)\mathbf{U} \rangle \geq 0$. Now if $\Upsilon(\mathbf{I}) = \|\Upsilon(\mathbf{I})\|_\infty \mathbf{I} = \alpha \mathbf{I}$, then pick

$$\mathbf{U} = \begin{bmatrix} -\Upsilon(\mathbf{a}) \\ \alpha \mathbf{I} \end{bmatrix}$$

Substitute $\mathbf{U}$ in the inner product above to get the result. This proves the Kadison–Schwarz inequality. □

Lemma 4.4.3. *Let $\mathcal{A}$ and $\mathcal{B}$ be von Neumann algebras of bounded linear operators on $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. A normal completely positive map $\Upsilon : \mathcal{A} \to \mathcal{B}$ can be written in the form*

$$\Upsilon(\mathbf{a}) = \mathbf{V}^* \pi(\mathbf{a}) \mathbf{V}, \quad \forall \mathbf{a} \in \mathcal{A},$$

where $\mathbf{V}$ is a bounded linear operator from $\mathbb{H}_B$ to a Hilbert space $\mathbb{H}_A$ and π is a normal representation of $\mathcal{A}$ in $\mathfrak{B}(\mathbb{H}_A)$.

Proof. Let $(\pi, \mathbf{V})$ be the minimal Stinespring representation of Υ with bounded linear operator $\mathbf{V} : \mathbb{H}_B \to \mathbb{H}_A$ and π is a *-homomorphism from $\mathcal{A}$ to $\mathfrak{B}(\mathbb{H}_A)$ such that $\Upsilon(\mathbf{a}) = \mathbf{V}^* \pi(\mathbf{a}) \mathbf{V}$ for all $\mathbf{a} \in \mathcal{A}$ (see Definition 4.3.2). We just need to check that the *-homomorphism $\pi : \mathcal{A} \to \mathfrak{B}(\mathbb{H}_A)$ is normal. Let $(\mathbf{x}_\alpha)_\alpha$ be a nondecreasing net of elements of $\mathcal{A}$ converging to $\mathbf{x} \in \mathcal{A}$ in the σ-weak topology. For all vectors $u, v \in \mathbb{H}_B$ and operators $\mathbf{a}, \mathbf{b} \in \mathcal{A}$, we have

$$\begin{aligned} &\lim_\alpha \langle \pi(\mathbf{b}) \mathbf{V} v, \pi(\mathbf{x}_\alpha) \pi(\mathbf{a}) \mathbf{V} u \rangle_{\mathbb{H}_A} \\ &\quad = \lim_\alpha \langle \mathbf{V} v, \pi(\mathbf{b})^* \pi(\mathbf{x}_\alpha) \pi(\mathbf{a}) \mathbf{V} u \rangle_{\mathbb{H}_A} \end{aligned}$$

$$= \lim_\alpha \langle \mathbf{V}v, \pi(\mathbf{b}^*\mathbf{x}_\alpha \mathbf{a})\mathbf{V}u\rangle_{\mathbb{H}_A} = \lim_\alpha \langle v, \mathbf{V}^*\pi(\mathbf{b}^*\mathbf{x}_\alpha \mathbf{a})\mathbf{V}u\rangle_{\mathbb{H}_A}$$
$$= \lim_\alpha \langle v, \Upsilon(\mathbf{b}^*\mathbf{x}_\alpha \mathbf{a})u\rangle_{\mathbb{H}_A} = \langle v, \Upsilon(\mathbf{b}^*\mathbf{x}\mathbf{a})u\rangle_{\mathbb{H}_A} \quad \text{(since } \Upsilon \text{ is normal)}$$
$$= \langle \pi(\mathbf{b})\mathbf{V}v, \pi(\mathbf{x})\pi(\mathbf{a})u\rangle_{\mathbb{H}_A}.$$

Thus, π is normal. □

Theorem 4.4.4 (Kraus theorem). *Let $\mathbb{H}$ be a separable complex Hilbert space, and let Υ be a *-weakly continuous linear map from $\mathfrak{B}(\mathbb{H})$ to $\mathfrak{B}(\mathbb{H})$. Then Υ is completely positive if and only if it is of the form*

$$\Upsilon(\mathbf{A}) = \sum_{n=1}^{+\infty} \Upsilon_n^* \mathbf{A} \Upsilon_n, \quad \forall \mathbf{A} \in \mathfrak{B}(\mathbb{H}), \tag{4.12}$$

for a sequence $(\Upsilon_n)_{n=1}^{+\infty}$ of bounded linear operators on $\mathbb{H}$ such that the series $\sum_{n=1}^{+\infty} \Upsilon_n^*\Upsilon_n$ is strongly convergent.

Proof. ($\Rightarrow$) Assume that Υ is completely positive. By the Stinespring's theorem (see Theorem 4.3.4), there exists a Stinespring representation $(\mathbb{K}, \pi, \mathbf{V})$ of Υ with π being normal. By Lemma 4.4.1, there exists a decomposition $\mathbb{K} = \bigoplus_{n=1}^{+\infty} \mathbb{K}_n$ such that each $\mathbb{K}_n$ is stable under π and such that π restricted to $\mathbb{K}_n$ is unitarily equivalent to the standard representation. Let $\mathbf{P}_n$ be the orthogonal projectors from $\mathbb{K}$ onto $\mathbb{K}_n$, let $\mathbf{U}_n : \mathbb{K}_n \to \mathbb{H}$ denote the unitary operator ensuring the equivalence. The operators $\mathbf{P}_n$ commute with the representation π. Hence,

$$\Upsilon(\mathbf{X}) = \mathbf{V}^*\pi(\mathbf{X})\mathbf{V} = \sum_{n=1}^{+\infty} \mathbf{V}^*\mathbf{P}_n\pi(\mathbf{X})\mathbf{P}_n\mathbf{V}$$
$$= \sum_{n=1}^{+\infty} \mathbf{V}^*\mathbf{P}_n\mathbf{U}_n\mathbf{X}\mathbf{U}_n^*\mathbf{P}_n\mathbf{V} = \sum_{n=1}^{+\infty} \Upsilon_n^*\mathbf{X}\Upsilon_n,$$

where $\Upsilon_n = \mathbf{U}_n^*\mathbf{P}_n\mathbf{V}$. Therefore, if $\Upsilon : \mathfrak{B}(\mathbb{H}) \to \mathfrak{B}(\mathbb{H})$ is completely positive, then Υ can be represented as (4.12).

($\Leftarrow$) Assume that Υ is of the form

$$\Upsilon(\mathbf{A}) = \sum_{n=1}^{+\infty} \Upsilon_n^* \mathbf{A} \Upsilon_n, \quad \forall \mathbf{A} \in \mathfrak{B}(\mathbb{H}), \tag{4.13}$$

for a sequence $(\Upsilon_n)_{n=1}^{+\infty}$ of bounded linear operators on $\mathbb{H}$ such that the series $\sum_{n=1}^{+\infty} \Upsilon_n^*\Upsilon_n$ is strongly convergent. By part 2 of Proposition 4.2.2, $\Upsilon(\mathbf{A})$ defined by the series in (4.13) converges strongly. By the Stinespring Theorem 4.3.1, $\Upsilon_n^*\mathbf{A}\Upsilon_n$ is completely positive for each n, so is $\Upsilon(\mathbf{A})$. This proves the theorem. □

5 Quantum channels and operations

Consider a quantum communication or information process scenario in which Alice, the sender, wishes to transmit quantum or classical information to Bob, the receiver, via a quantum device. This chapter explores the mathematical theory of basic ingredients in quantum communication and information processing including the notion of quantum channels that describes the quantum devices through which storage or transmission of information takes place. Any quantum information process, be it storage or transfer, can be represented as a quantum channel. Roughly speaking, a quantum channel is defined as a specific type of completely positive trace-preserving operator from the set of states of the input quantum system $\mathbb{H}_A$ into the set of states of the output quantum system $\mathbb{H}_B$. Such operators describe discrete time evolution of an open quantum system interacting with an environment.

In the quantum information theory literature, the vast majority of attention up to now was paid to study of finite-dimensional memoryless systems (see, e. g., popular quantum information theory books by Holevo [77], Hayashi [61], Watrous [173] and Wilde [178] and references given therein). However, interest in infinite-dimensional systems has been increasing in recent years. Many aspects of quantum information theory have been recently extended by M. E. Shirokov, A. S. Holevo and their collaborators to infinite dimensions. Without explicitly listing their contributions, readers are referred to their works listed in the bibliography section at the end of this book.

This chapter systematically explores the concept and properties of quantum channels and quantum operations in infinite dimensions.

5.1 Quantum operations

Quantum operations are formulated in terms of the quantum state or density operator description of a quantum mechanical system. Rigorously, a quantum operation is a linear, completely positive and trace-nondecreasing map from the set of quantum states of one system into that of another.

To define quantum operations, we recall from Definition 1.8.2 that $\mathfrak{T}(\mathbb{H})$ is the Banach space of trace-class operators $\mathbf{T}$ on the complex Hilbert space $\mathbb{H}$ under the trace-norm $\|\mathbf{T}\|_1$ defined by $\|\mathbf{T}\|_1 := \mathrm{tr}[|\mathbf{T}|] = \mathrm{tr}[\sqrt{\mathbf{T}^*\mathbf{T}}]$.

A linear map $\Phi : \mathfrak{T}(\mathbb{H}_A) \to \mathfrak{T}(\mathbb{H}_B)$ is said to be trace nonincreasing if

$$\mathrm{tr}[\Phi(\rho)] \le \mathrm{tr}[\rho], \quad \forall \rho \in \mathfrak{T}(\mathbb{H}_A), \tag{5.1}$$

and $\Phi : \mathfrak{T}(\mathbb{H}_A) \to \mathfrak{T}(\mathbb{H}_B)$ is said to be trace preserving if

$$\mathrm{tr}[\Phi(\rho)] = \mathrm{tr}[\rho], \quad \forall \rho \in \mathfrak{T}(\mathbb{H}_A). \tag{5.2}$$

https://doi.org/10.1515/9783110788105-005

Definition 5.1.1. A linear completely positive trace-nonincreasing map

$$\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$$

is called a quantum operation from system A to system B.

Definition 5.1.2. A linear completely positive trace-nonincreasing map

$$\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$$

is called an extended quantum operation from system A to system B.

The collection of quantum operations from A to B will be denoted by $\mathfrak{QO}(A,B)$ and the collection of extended quantum operations from system A to system B will be denoted by $\mathfrak{EQO}(A,B)$. When $\mathbb{H}_A = \mathbb{H}_B$, $\mathfrak{QO}(A,B)$ and $\mathfrak{EQO}(A,B)$ will be written as $\mathfrak{QO}(A)$ and $\mathfrak{EQO}(A)$, respectively.

Trace-preserving and trace-nonincreasing positive linear maps between $\mathfrak{T}_+(\mathbb{H}_A)$ and $\mathfrak{T}_+(\mathbb{H}_B)$ can be considered noncommutative analogs of Markov and sub-Markov maps in the classical probability theory.

5.2 Quantum channels

Roughly speaking, a quantum channel is a specific type of map from one quantum system to another, which will be described mathematically in Schrodinger and Heisenberg pictures in the next two subsections. In the Schrodinger picture, the quantum channels will be presented as being the resulting transformation of a quantum state of $\mathbb{H}_A$ after a contact and an evolution with some environment to a quantum state of $\mathbb{H}_B$. As usual, for all quantum evolutions there is a dual picture, an Heisenberg picture, where the evolution is seen from the point of view of observables instead of states.

To explore the relationship between the Shrodinger picture and Heisenberg picture of a quantum system represented by a separable Hilbert space $\mathbb{H}$, consider the Banach spaces of trace-class operators $\mathfrak{T}(\mathbb{H})$ under trace-norm $\|\cdot\|_1$ and the Banach space of bounded linear operators $\mathfrak{B}(\mathbb{H})$ under the operator norm $\|\cdot\|_\infty$. It has been shown in Proposition 2.3.14 that $\mathfrak{T}(\mathbb{H})$ is a predual of $\mathfrak{B}(\mathbb{H})$ via the bilinear relation $\langle\langle \cdot,\cdot \rangle\rangle : \mathfrak{B}(\mathbb{H}) \times \mathfrak{T}(\mathbb{H}) \to \mathbb{C}$ defined by

$$\langle\langle \mathbf{a}, \mathbf{T} \rangle\rangle = \operatorname{tr}[\mathbf{aT}], \quad \forall \mathbf{a} \in \mathfrak{B}(\mathbb{H}) \text{ and } \forall \mathbf{T} \in \mathfrak{T}(\mathbb{H}).$$

In other words, $\mathfrak{T}^*(\mathbb{H})$, the topological dual of $\mathfrak{T}(\mathbb{H})$, equals $\mathfrak{B}(\mathbb{H})$. Moreover, for any $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$, its operator norm $\|\mathbf{T}\|_\infty$ is related to its trace-class norm $\|\mathbf{T}\|_1$ via the following relation:

$$\|\mathbf{T}\|_\infty = \sup_{\|\rho\|_1 \le 1} \|\mathbf{T}\rho\|_1. \tag{5.3}$$

In particular, if $\mathbf{T}$ is such that $\|\mathbf{T}\|_1 = 1$, then $\|\mathbf{T}\|_\infty = \|\mathbf{T}\|_1 = 1$.

5.2.1 Quantum channels in the Schrodinger picture

Let $\mathcal{S}(\mathbb{H}_A)$ and $\mathcal{S}(\mathbb{H}_B)$ be the collections of quantum states on $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively.

The definition and basic properties of quantum channels in Schrodinger picture are given below.

Definition 5.2.1. A linear completely positive trace-preserving map

$\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$, $\rho_B = \Phi(\rho_A)$, is called a quantum channel from system A to system B. In this case, $\rho_A \in \mathcal{S}(\mathbb{H}_A)$ will be called an input state and $\rho_B \in \mathcal{S}(\mathbb{H}_B)$ will be called the output state of the channel.

The collection of quantum channels from system A to system B will be denoted by $\mathfrak{QC}(A, B)$. When $\mathbb{H}_A = \mathbb{H}_B$, we will simply write $\mathfrak{QC}(A, B)$ as $\mathfrak{QC}(A)$.

In functional analysis, a partial isometry (see also Theorem 1.8.11 for a description) is a linear map Υ between Hilbert spaces $\mathbb{H}$ and $\mathbb{K}$ such that Υ is an isometry between $(\ker(\Upsilon))^\perp \subset \mathbb{H}$ and $\text{range}(\Upsilon) \subset \mathbb{K}$, where $(\ker(\Upsilon))^\perp$ (the orthogonal complement of its kernel) is called the initial subspace and $\text{range}(\Upsilon)$ (the range of Υ) is called the final subspace of the map. The concept of partial isometry can be defined in other equivalent ways. If $\mathbf{U}$ is an isometric map defined on a closed subset $\mathbb{H}_0$ of a Hilbert space $\mathbb{H}$, then we can define an extension $\mathbf{W}$ of $\mathbf{U}$ to all of $\mathbb{H}$ by the condition that $\mathbf{W}$ be zero on $\mathbb{H}_0^\perp$ (the orthogonal complement of $\mathbb{H}_0$). Thus, a partial isometry is also sometimes defined as a closed partially defined isometric map.

Definition 5.2.2 (Isometrical equivalence). Let A, B and B' be quantum systems represented by the separable complex Hilbert spaces $\mathbb{H}_A$, $\mathbb{H}_B$ and $\mathbb{H}_{B'}$, respectively. The (extended) quantum channels $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ and $\Phi' : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_{B'})$ are said to be *isometrically equivalent* if there exists a *partial isometry* $\mathbf{W} : \mathbb{H}_B \to \mathbb{H}_{B'}$ such that

$$\Phi'(\rho) = \mathbf{W}\Phi(\rho)\mathbf{W}^*, \quad \Phi(\rho) = \mathbf{W}^*\Phi'(\rho)\mathbf{W}, \quad \forall \rho \in \mathfrak{T}(\mathbb{H}_A). \tag{5.4}$$

The notion of isometrical equivalence is very close to the notion of unitary equivalence. Indeed, the isometrical equivalence of the channels Φ and Φ' means unitary equivalence of these channels with the output spaces $\mathbb{H}_B$ and $\mathbb{H}_{B'}$ replaced by their subspaces $\mathbb{H}_B^\Phi = \vee_{\rho \in \mathcal{S}(\mathbb{H}_A)} \text{supp}(\Phi(\rho))$ and $\mathbb{H}_{B'}^{\Phi'} = \vee_{\rho \in \mathcal{S}(\mathbb{H}_A)} \text{supp}(\Phi'(\rho))$. We use the notion of isometrical equivalence, since dealing with a given representation of a quantum channel Φ it not easy in general to determine the corresponding subspace $\mathbb{H}_B^\Phi$.

Some examples of quantum channels are given below.

Example 5.1 (Isometric channel). Let $\Upsilon : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be an isometric mapping and let $\Upsilon^* : \mathfrak{B}(\mathcal{H}_B) \to \mathfrak{B}(\mathcal{H}_A)$ be its adjoint operator. We claim that the map $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ defined by

$$\Phi(\rho) = \Upsilon\rho\Upsilon^*, \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A), \tag{5.5}$$

is a quantum channel. To prove this claim, we first note that $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is completely positive. This is because for $u_1, u_2, \dots, u_n \in \mathbb{H}_A$ and $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathfrak{B}(\mathbb{H}_B)$, we have

$$\begin{aligned}\sum_{1\le i,j\le n} \langle u_i, \Phi^*(\mathbf{a}_i^*\mathbf{a}_j)u_j\rangle_{\mathbb{H}_A} &= \sum_{1\le i,j\le n} \langle u_i, \Upsilon\mathbf{a}_i\mathbf{a}_j^*\Upsilon^* u_j\rangle_{\mathbb{H}_A} \\ &= \sum_{1\le i,j\le n} \langle \Upsilon\mathbf{a}_j u_j, \Upsilon\mathbf{a}_i u_i\rangle_{\mathbb{H}_A} \ge \sum_{i=1}^n \langle \Upsilon\mathbf{a}_i u_i, \Upsilon\mathbf{a}_i u_i\rangle_{\mathbb{H}_A} = \sum_{i=1}^n \|\Upsilon\mathbf{a}_i u_i\|^2_{\mathbb{H}_A} \\ &\ge 0.\end{aligned}$$

By Proposition 4.1.6, $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ is completely positive. Hence, $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is completely positive. Second, we note that Φ is trace preserving. This is because

$$\operatorname{tr}[\Phi(\rho)] = \operatorname{tr}[\Upsilon\rho\Upsilon^*] = \operatorname{tr}[\Upsilon\Upsilon^*\rho] = \operatorname{tr}[\rho].$$

Therefore, Φ is a quantum channel. The channel defined by (5.5) will be referred to as an isometric channel. If $\mathbb{H}_A = \mathbb{H}_B = \mathbb{H}$, then $\mathbf{U}$ is a unitary operator and the channel $\Phi : \mathcal{S}(\mathbb{H}) \to \mathcal{S}(\mathbb{H})$ is called a unitary channel.

Since $\mathcal{S}(\mathbb{H}_A)$ is a closed convex subset of $\mathfrak{T}_+(\mathbb{H}_A)$, the quantum channel Φ can be easily extended from $\mathcal{S}(\mathbb{H}_A)$ to $\mathfrak{T}_+(\mathbb{H}_A)$ by linear extension. In this case, the extended map $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ is also a linear completely positive map, which may not preserve its trace in general (see Shirokov [155]).

We denote the class of quantum channels (resp., extended quantum channels) from system A to system B by $\mathfrak{QC}(A,B)$ (resp., $\mathfrak{EQC}(A,B)$). As usual, we write $\mathfrak{QC}(A,B)$ as $\mathfrak{QC}(A)$ and $\mathfrak{EQC}(A,B)$ as $\mathfrak{EQC}(A)$ when $A = B$.

5.2.2 Quantum channels in the Heisenberg picture

Without loss generality, we can and often consider the extended version of channel $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ in place of the channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ itself. In this case, Φ^*, the adjoint of the (extended) quantum channel Φ is a map from $\mathfrak{B}(\mathbb{H}_B)$ to $\mathfrak{B}(\mathbb{H}_A)$. This is because $\mathfrak{T}(\mathbb{H}_A)$ and $\mathfrak{T}(\mathbb{H}_B)$ are preduals of $\mathfrak{B}(\mathbb{H}_A)$ and $\mathfrak{B}(\mathbb{H}_B)$, respectively.

Definition 5.2.3. Let $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ be an (extended) quantum channel from A to B. The adjoint operator of Φ, $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ is called the dual channel of Φ.

The class of dual channels is denoted by $\mathfrak{EQC}^*(B, A)$, and similarly $\mathfrak{EQC}^*(B, A)$ as $\mathfrak{EQC}^*(A)$ if $A = B$.

The (extended) quantum channels $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ implicitly assume the Schrodinger picture in which the states of the system are evolved while the observables of the system are kept fixed. On the other hand, the dual channel $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ is in the Heisenberg picture and it implies that the states of the system are fixed and the observables evolve in time. The (extended) quantum channel Φ and its associated dual channel Φ^*, however, satisfy the following *duality relation*:

$$\mathrm{tr}[\Phi(\rho_A)\mathbf{O}_B] = \mathrm{tr}[\rho_A \Phi^*(\mathbf{O}_B)], \quad \forall \rho_A \in \mathfrak{T}(\mathbb{H}_A) \text{ and } \forall \mathbf{O}_B \in \mathfrak{B}(\mathbb{H}_B). \tag{5.6}$$

Notice also that in the duality relation the concatenation of channels goes in reversed order in the Schrodinger picture. That is, given (extended) channels $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ and $\Psi : \mathfrak{T}_+(\mathbb{H}_B) \to \mathfrak{T}_+(\mathbb{H}_C)$, we have

$$(\Psi \circ \Phi)^* = \Phi^* \circ \Psi^*.$$

A physical interpretation of the dual quantum channel $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ is the following: when the system is initially in the state $\rho \in \mathfrak{T}_+(\mathbb{H}_A)$, the expectation value of the measurement of the observable $\mathbf{B} \in \mathfrak{B}(\mathbb{H}_B)$ at the output side of the channel is given in terms of Φ by $\mathrm{tr}[\rho\Phi^*(\mathbf{B})]$.

The following result characterizes the dual channel Φ^* of Φ, which can be viewed as the quantum channel Φ in the Heisenberg picture.

Proposition 5.2.4. *The dual channel* $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ *of the (extended) quantum channel* $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ *is a unital completely positive map.*

Proof. The dual channel Φ^* is completely positive because the topological dual of a completely positive map is completely positive but not necessarily trace preserving. However, Φ^* is unital, i. e., $\Phi^*(\mathbf{I}_B) = \mathbf{I}_A$. This is because for all $\rho \in \mathfrak{T}_+(\mathbb{H}_A)$,

$$\mathrm{tr}[\rho] = \langle\langle \rho, \mathbf{I}_A \rangle\rangle = \mathrm{tr}[\rho \mathbf{I}_A] = \mathrm{tr}[\Phi(\rho)\mathbf{I}_B] = \mathrm{tr}[\rho\Phi^*(\mathbf{I}_B)] = \langle\langle \rho, \Phi^*(\mathbf{I}_B) \rangle\rangle.$$

This implies that $\Phi^*(\mathbf{I}_B) = \mathbf{I}_A$. This proves the proposition. □

Example 5.2. In this example, we briefly visit a special case of quantum channels $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ in the Heisenberg picture that converts classical information to a certain type of quantum information. A channel in the Heisenberg picture converting classical information to a type of quantum information is a channel $\Phi^* : \mathcal{B} \to \mathcal{C}(\mathbb{X})$

with some finite set $\mathbb{X} = \{x_1, x_2, \dots, x_{|\mathbb{X}|}\}$ (the symbol $|X|$ denotes the cardinality of or the number of elements in the set X) that represents a codeword of classical information, where $\mathcal{B} \subset \mathfrak{B}(\mathbb{H}_B)$ is a C^*-algebra of bounded linear operators on $\mathbb{H}_B$ and $C(\mathbb{X})$ denotes the class of continuous real-valued functions defined on $\mathbb{X}$. To interpret $\mathcal{C}(\mathbb{X})$ as an algebra of operators acting on a Hilbert space $\mathbb{H}_A$, we choose an arbitrary but fixed orthonormal basis $|x\rangle$, $(x \in \mathbb{X})$ in $\mathbb{H}_B$ and identify the function $f \in \mathcal{C}(\mathbb{X})$ with the operator $f = \sum_x f_x |x\rangle\langle x| \in \mathfrak{B}(\mathbb{H}_B)$ (we use the same symbol for the function and the operator, provided confusion can be avoided). Most frequently we can choose $\mathbb{H}_B = \mathbb{C}_{|\mathbb{X}|\times\mathbb{X}|}$ and the canonical basis for $|x\rangle$. Hence, $\mathcal{C}(\mathbb{X})$ becomes the algebra of diagonal $|\mathbb{X}| \times |\mathbb{X}|$ matrices. Since $\mathcal{C}(\mathbb{X})$ is finite-dimensional and admits the distinguished basis $|x\rangle\langle x|$, $x \in \mathbb{X}$, and it is naturally isomorphic to its dual $\mathcal{C}^*(\mathbb{X})$. The channel Φ^* can be given by the decomposition

$$\Phi^*(\mathbf{B}) = \sum_{i=1}^{|\mathbb{X}|} \rho_{x_i}(\mathbf{B}) e_{x_i}, \quad \forall \mathbf{B} \in \mathcal{B}, \tag{5.7}$$

where for each $i = 1, 2, \dots, |\mathbb{X}|$, $\rho_{x_i} : \mathbf{B} \mapsto \rho_{x_i}(\mathbf{B}) := \Phi^*(\mathbf{B})(x_i)$ is a positive normalized functional on $\mathfrak{B}(\mathbb{H}_B)$, i. e., a state in the sense of Definition 2.4.1. Hence, a channel of this type describes a parameter-dependent preparation, or preparator. If in addition the output system is likewise classical, i. e., $\Phi^* : \mathcal{C}(\mathbb{X}) \to \mathcal{C}(\mathbb{X})$, in this case the channel Φ^* is completely specified by the $(|\mathbb{X}| \times |\mathbb{X}|)$ matrix $(T_{ij})_{i,j=1}^{|\mathbb{X}|}$ with

$$T_{ij} := \Phi^*_{x_i x_j} = \Phi^*(e_{x_j}) x_i, \quad \forall i, j = 1, 2, \dots, |\mathbb{X}|,$$

describing the probability to receive the symbol x_j when the symbol x_i was sent. In terms of the transition matrix $[\Phi^*_{ij}]_{i,j=1}^{|\mathbb{X}|}$, equation (5.7) can be rewritten as follows:

$$(\Phi^* f)(x_i) = \sum_{j=1}^{|\mathbb{X}|} T_{ij} f_{x_j} \tag{5.8}$$

for any $f \in \mathcal{C}(\mathbb{X})$, where $f_{x_j} = f(x_j)$ and $j = 1, \dots, |\mathbb{X}|$. Dually, a measurement is simply a channel $\mathbf{T} : \mathcal{A} \to \mathcal{B}$ with classical output algebra $\mathcal{A} = C(\mathbb{X})$, for some finite set $\mathbb{X}$. Then $\mathbf{T}$ is completely specified by its values $\mathbf{E}_x := \mathbf{T}(e_x)$ on the basis $\{e_x\}_{x=1}^{|\mathbb{X}|}$ of $\mathcal{A}$, via $x = 1$,

$$\mathbf{T}(f) = \sum_{x=1}^{|\mathbb{X}|} f_x \mathbf{E}_x,$$

and any such map $\mathbf{T}$ is a channel if and only if the maps $\mathbf{E}_x : A \to B$ are positive and $\{\mathbf{E}_x\}_{x=1}^{|\mathbb{X}|}$ forms a basis of $\mathcal{B}$.

5.3 Representations and dilations of channels

5.3.1 Stinespring representation

Recall that if $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ is an extended quantum channel from system A to system B, then its dual channel $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ is a unital (i. e., $\Phi^*(\mathbf{I}_B) = \mathbf{I}_A$) completely positive map (see Proposition 4.2.2). Application of the Stinespring theorem (see Theorem 4.3.1 for Stinespring representation of completely positive maps) to the dual channel Φ^* yields existence of an environmental quantum system E represented by the Hilbert space $\mathbb{H}_E$ and of an isometric (i. e., $\mathbf{V}^*\mathbf{V} = \mathbf{I}_A$) map $\mathbf{V} : \mathbb{H}_A \to \mathbb{H}_B \otimes \mathbb{H}_E$, such that Φ^* can be written as

$$\Phi^*(\mathbf{B}) = \mathbf{V}^*(\mathbf{B} \otimes \mathbf{I}_E)\mathbf{V}, \quad \forall \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B), \tag{5.9}$$

where $\mathbf{I}_E$ is the identity operator on $\mathbb{H}_E$.

The duality relation $\operatorname{tr}[\Phi(\rho)\mathbf{B}] = \operatorname{tr}[\rho\Phi^*(\mathbf{B})]$ between Φ and Φ^* yields

$$\begin{aligned}
\operatorname{tr}[\Phi(\rho)\mathbf{B}] &= \operatorname{tr}[\rho\Phi^*(\mathbf{B})] = \operatorname{tr}[\rho\mathbf{V}^*(\mathbf{B} \otimes \mathbf{I}_E)\mathbf{V}] \\
&= \operatorname{tr}[\mathbf{V}^*(\mathbf{B} \otimes \mathbf{I}_E)\mathbf{V}\rho] \quad \text{(by Part 2 of Proposition 1.8.4)} \\
&= \operatorname{tr}[\operatorname{tr}_E[\mathbf{V}^*(\mathbf{B} \otimes \mathbf{I}_E)\mathbf{V}\rho]] = \operatorname{tr}[\operatorname{tr}_E[\mathbf{V}\rho\mathbf{V}^*(\mathbf{B} \otimes \mathbf{I}_E)]] \\
&= \operatorname{tr}[\operatorname{tr}_E[\mathbf{V}\rho\mathbf{V}^*]\mathbf{B}], \quad \forall \rho \in \mathfrak{T}(\mathbb{H}_A) \text{ and } \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B).
\end{aligned}$$

Therefore, the quantum channel $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ have the following representations in Schrodinger picture:

$$\Phi(\rho) = \operatorname{tr}_E[\mathbf{V}\rho\mathbf{V}^*], \quad \forall \rho \in \mathfrak{T}_+(\mathbb{H}_A). \tag{5.10}$$

Either one of the two representations, (5.10) of Φ and (5.9) of Φ^*, will be called a Stinespring representation and be denoted by $(\mathbb{H}_E, \mathbf{V})$. The Stinespring representation $(\mathbb{H}_E, \mathbf{V})$ is called minimal if the subspace

$$\mathcal{M} = \{(\mathbf{B} \otimes \mathbf{I}_E)\mathbf{V}|\varphi\rangle_A \mid \varphi \in \mathbb{H}_A, \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B)\} \tag{5.11}$$

is dense in $\mathbb{H}_{BE}$. Stinespring's representation is not at all unique. In fact, if $(\mathbb{H}_E, \mathbf{V})$ is a representation for the channel Φ^*, then it is easily seen that a further representation for Φ^* is given by $(\mathbb{H}_E, (\mathbf{I}_B \otimes \mathbf{U})\mathbf{V})$ with any unitary $\mathbf{U} \in \mathfrak{B}(\mathbb{H}_B)$. However, it is straightforward to show that the minimal Stinespring representation is unique up to such unitary equivalence: Assume that the dual quantum channel Φ^* has a minimal Stinespring representation $(\mathbb{H}_E, \mathbf{V})$ as well as a not necessary minimal one $(\mathbb{H}_{\tilde{E}}, \tilde{\mathbf{V}})$. That is,

$$\Phi^*(\mathbf{B}) = \tilde{\mathbf{V}}^*(\mathbf{B} \otimes \mathbf{I}_{\tilde{E}})\tilde{\mathbf{V}}, \quad \forall \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B)$$

for an isometric map $\tilde{\mathbf{V}} : \mathbb{H}_A \to \mathbb{H}_{B\tilde{E}}$ and for a possibly different dilation space $\mathbb{H}_{\tilde{E}}$. Since $(\mathbb{H}_E, \mathbf{V})$ is chosen to be minimal, we can conclude that $\dim(\mathbb{H}_E) \leq \dim(\mathbb{H}_{\tilde{E}})$. Define the map $\tilde{\mathbf{U}} : \mathbb{H}_{BE} \to \mathbb{H}_{B\tilde{E}}$ by setting

$$\tilde{\mathbf{U}}(\mathbf{B} \otimes \mathbf{I}_E)\mathbf{V}\varphi := (\mathbf{B} \otimes \mathbf{I}_{\tilde{E}})\tilde{\mathbf{V}}\varphi, \quad \forall \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B) \text{ and } \forall \varphi \in \mathbb{H}_A. \tag{5.12}$$

It can be easily shown that the map $\tilde{\mathbf{U}}$ defined above is a well-defined isometry. In particular, by choosing $\mathbf{B} = \mathbf{I}_B$ in (5.12), we see $\tilde{\mathbf{U}}\mathbf{V} = \tilde{\mathbf{V}}$. From the definition of $\tilde{\mathbf{U}}$, we immediately have the following interwinding relation:

$$\tilde{\mathbf{U}}(\mathbf{B} \otimes \mathbf{I}_E) = (\mathbf{B} \otimes \mathbf{I}_{\tilde{E}})\tilde{\mathbf{U}}, \quad \forall \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B). \tag{5.13}$$

Hence, $\tilde{\mathbf{U}}$ must be decomposable as $\tilde{\mathbf{U}} = \mathbf{B} \otimes \mathbf{U}$ for some isometry $\mathbf{U} : \mathbb{H}_E \to \mathbb{H}_{\tilde{E}}$. If both representations $(\mathbb{H}_E, \mathbf{V})$ and $(\mathbb{H}_{\tilde{E}}, \tilde{\mathbf{V}})$ are minimal, then $\dim(\mathbb{H}_E) = \dim(\mathbb{H}_{\tilde{E}})$, and $\mathbf{U}$ is unitary as suggested.

We offer a physical interpretation of Stinespring's representation as follows. The dilation space $\mathbb{H}_E$ represents the environment and the Stinespring's isometry $\mathbf{V}$ transforms the input state ρ into the state $\mathbf{V}\rho\mathbf{V}^*$ on $\mathbb{H}_{BE}$, which is correlated between the output and the environment. The output state $\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}\rho\mathbf{V}^*] \in \mathfrak{T}(\mathbb{H}_B)$ is then obtained by tracing out $\mathbf{V}\rho\mathbf{V}^*$ over the environment $\mathbb{H}_E$. Physically, one would expect in the above Stinespring representation of the channel Φ (or its dual channel Φ^*), a unitary operation $\mathbf{U}$ instead of an isometric $\mathbf{V}$. However, the initial state of the environment in this case can be considered fixed, effectively reducing $\mathbf{U}$ to an isometry, $\mathbf{V}\psi := \mathbf{U}(\psi \otimes \sigma_0)$ for some fixed initial pure state $|\sigma_0\rangle_{\mathbb{E}}$ of the environment system.

5.3.2 Unitary dilation

This subsection explores different type of dilations of quantum channels as completely positive and trace-preserving maps from $\mathfrak{T}_+(\mathbb{H}_A)$ to $\mathfrak{T}_+(\mathbb{H}_B)$ in the Schrodinger picture and its dual channels as unital completely positive maps from $\mathfrak{B}(\mathbb{H}_B)$ to $\mathfrak{B}(\mathbb{H}_A)$ in the Heisenberg picture.

In the following, we explore unitary dilation (a special case of Stinespring dilation) and Kraus representation for quantum channels that are described in the Schrodinger picture.

The general version of unitary dilation of (extended) quantum channels from system A to system B holds (see Remark 5.2 below). Unfortunately, the proof is rather involved and only the case when $\mathbb{H}_A = \mathbb{H}_B = \mathbb{H}$ will be illustrated in the following result (see, e. g., Attal [3]).

Theorem 5.3.1 (Unitary dilation). *Let $\mathbb{H}$ be a complex separable Hilbert space and let Φ be quantum channel from $\mathbb{H}$ to $\mathbb{H}$. Then there exists a complex separable Hilbert space $\mathbb{K}$, a quantum state ω on $\mathbb{K}$ and a unitary operator $\mathbf{U}$ on $\mathbb{H} \otimes \mathbb{K}$ such that*

$$\Phi(\mathbf{T}) = \mathrm{tr}_{\mathbb{K}}\big[\mathbf{U}(\mathbf{T}\otimes\omega)\mathbf{U}^*\big], \quad \forall \mathbf{T} \in \mathfrak{T}_+(\mathbb{H}). \tag{5.14}$$

Furthermore, it is always possible to choose $\mathbb{K}$, $\mathbf{U}$ *and* ω *in such a way that* ω *is a pure state.*

Proof. Let $\Phi : \mathfrak{T}_+(\mathbb{H}) \to \mathfrak{T}_+(\mathbb{H})$ be an extended quantum channel. Then the Kraus theorem (see Theorem 4.4.4) implies that Φ can be expressed in the form

$$\Phi(\mathbf{T}) = \sum_{i\in\mathbb{I}} \mathbf{M}_i\mathbf{T}\mathbf{M}_i^*,$$

in the strong convergence sense, where the $\mathbf{M}_i$'s are bounded linear operators on $\mathbb{H}$ and $\sum_{i\in\mathbb{I}}\mathbf{M}_i^*\mathbf{M}_i = \mathbf{I}_{\mathbb{H}}$. We consider the Hilbert space $\mathbb{K}$ with some orthonormal basis $\{\psi_i\}_{i\in\mathbb{I}}$ indexed by the same set $\mathbb{I}$ that contains 0 (which may be finite or countably infinite). Define the linear operator $\mathbf{V} : \mathbb{H}\otimes\mathbb{C}|\psi_0\rangle_{\mathbb{K}} \to \mathbb{H}\times\mathbb{K}$,

$$\mathbf{V}(\phi\otimes\psi_0) = \sum_{j\in\mathbb{I}}(\mathbf{M}_j\phi)\otimes\psi_0.$$

We claim that $\mathbf{V}$ is an isometry. To prove the claim, we use Proposition 4.2.3 and note that

$$\begin{aligned}
\|\mathbf{V}(\psi\otimes\psi_0)\|^2_{\mathbb{H}\otimes\mathbb{K}} &= \Big\langle \sum_{i\in\mathbb{I}}(\mathbf{M}_i\phi)\otimes\psi_0, \sum_{j\in\mathbb{I}}(\mathbf{M}_j\phi)\otimes\psi_0 \Big\rangle_{\mathbb{H}\otimes\mathbb{K}} \\
&= \sum_{j\in\mathbb{I}}\|\mathbf{M}_j\phi\|^2_{\mathbb{H}} = \Big\langle \sum_{j\in\mathbb{I}}\mathbf{M}_j\phi, \sum_{j\in\mathbb{I}}\mathbf{M}_j\phi \Big\rangle_{\mathbb{H}} \\
&= \Big\langle \phi, \sum_{j\in\mathbb{I}}\mathbf{M}_j^*\mathbf{M}_j\phi \Big\rangle_{\mathbb{H}} = \|\phi\|^2_{\mathbb{H}} = \|\phi\otimes\psi_0\|^2_{\mathbb{H}\otimes\mathbb{K}}.
\end{aligned}$$

We now wish to extend the operator $\mathbf{V}$ into a unitary operator from $\mathbb{H}\otimes\mathbb{K}$ to $\mathbb{H}\otimes\mathbb{K}$. We consider following cases:

Case (1): Assume first that $\mathbb{H}$ is finite-dimensional, then the range of $\mathbf{V}$ is a finite-dimensional subspace of $\mathbb{H}\otimes\mathbb{K}$, with same dimension as $\mathbb{H}$. If $\mathbb{K}$ is finite-dimensional, then $\mathbb{H}\otimes(\mathbb{K}\ominus\mathbb{C}|\psi_0\rangle_{\mathbb{K}})$ (where $\mathbb{K}\ominus\mathbb{C}|\psi_0\rangle_{\mathbb{K}}$ is the subspace of $\mathbb{K}$ such that $(\mathbb{K}\ominus\mathbb{C}|\psi_0\rangle_{\mathbb{K}})\oplus\mathbb{C}|\psi\rangle_{\mathbb{K}} = \mathbb{K}$) is of the same dimension as $(\mathrm{range}(\mathbf{V}))^\perp$; if $\mathbb{K}$ is infinite-dimensional, then $\mathbb{H}\otimes(\mathbb{K}\ominus\mathbb{C}|\psi_0\rangle_{\mathbb{K}})$ and $(\mathrm{range}(\mathbf{V}))^\perp$ are both infinite-dimensional separable Hilbert spaces. In both cases, the space $(\mathrm{range}(\mathbf{V}))^\perp$ can be mapped unitarily to $\mathbb{H}\otimes(\mathbb{K}\ominus\mathbb{C}|\psi_0\rangle_{\mathbb{K}})$ through a unitary operator $\mathbf{W}$. The operator $\hat{\mathbf{V}}$ from $\mathbb{H}\otimes\mathbb{K}$ to itself, which acts as $\mathbf{V}$ on $\mathbb{H}\otimes\mathbb{C}|\psi_0\rangle_{\mathbb{K}}$ and as $\mathbf{W}^*$ on $\mathbb{H}\otimes(\mathbb{K}\ominus\mathbb{C}|\psi_0\rangle_{\mathbb{K}})$ is then unitary and extends $\mathbf{V}$.

Case (2): If $\mathbb{H}$ is infinite-dimensional, it may happen that $\mathrm{range}(\mathbf{V})$ is the whole of $\mathbb{H}\otimes\mathbb{K}$, so we cannot directly extend $\mathbf{V}$ into a unitary operator on $\mathbb{H}\otimes\mathbb{K}$. We embed $\mathbb{K}$ into a larger Hilbert space $\mathbb{K}'$ by adding one new vector, ψ_{-1} say, orthogonal to all $\mathbb{K}$. The space $(\mathrm{range}(\mathbf{V}))^\perp$ in $\mathbb{H}\otimes\mathbb{K}'$ is then infinite-dimensional (separable), as is $\mathbb{H}\otimes$

$(\mathbb{K}' \ominus \mathbb{C}|\psi_0\rangle_{\mathbb{K}})$. Once again one can unitarily map $\mathbb{H} \otimes (\mathbb{K}' \ominus \mathbb{C}|\psi_0\rangle_{\mathbb{K}})$ onto $(\mathrm{range}(\mathbf{V}))^{\perp}$ and this way obtain an extension of $\mathbf{V}$ into a unitary operator $\hat{\mathbf{V}}$ from $\mathbb{H}\otimes\mathbb{K}'$ onto itself.

We conclude from the above that we have obtained: (i) a Hilbert space, which we denote by $\mathbb{K}$, with an orthonormal basis $\{\psi_j\}_{j\in\mathbb{J}}$ where the index set $\mathbb{J}$ contains the original set $\mathbb{I}$ and (ii) a unitary extension $\hat{\mathbf{V}}$ of $\mathbf{V}$ from $\mathbb{H} \otimes \mathbb{K}$ onto itself.

We want now to prove that they provide the announced quantum channel. Let $\mathbf{T}$ be a trace-class operator, which we first assume to be a pure state $|\varphi\rangle_{\mathbb{H}}\langle\varphi|$. If we compute

$$\mathrm{tr}_{\mathbb{K}}[\hat{\mathbf{V}}(\mathbf{T} \otimes |\psi_0\rangle_{\mathbb{K}}\langle\psi|_0)\hat{\mathbf{V}}^*],$$

we obtain, by considering the orthonormal basis $\{\psi_j\}_{j\in\mathbb{J}}$ of $\mathbb{K}$ as described above,

$$\begin{aligned}
\mathrm{tr}_{\mathbb{K}}[\hat{\mathbf{V}}(\mathbf{T} \otimes |\psi_0\rangle_{\mathbb{K}}\langle\psi_0|)\hat{\mathbf{V}}^*] &= \sum_{i\in\mathbb{J}} \langle\psi_i, \hat{\mathbf{V}}(\mathbf{T} \otimes |\psi_0\rangle_{\mathbb{K}}\langle\psi_0|)\hat{\mathbf{V}}^*\psi_i\rangle_{\mathbb{K}} \\
&= \sum_{i\in\mathbb{J}} \langle\psi_i, \hat{\mathbf{V}}(|\varphi\rangle_{\mathbb{H}}\langle\varphi| \otimes |\psi_0\rangle_{\mathbb{K}}\langle\psi_0|)\hat{\mathbf{V}}^*\psi_i\rangle_{\mathbb{K}} \\
&= \sum_{i\in\mathbb{J}} \langle\psi_i, \hat{\mathbf{V}}(|\varphi \otimes |\psi_0\rangle_{\mathbb{H}\otimes\mathbb{K}}\langle\varphi \otimes \psi_0|)\hat{\mathbf{V}}^*\psi_i\rangle_{\mathbb{K}} \\
&= \sum_{i\in\mathbb{J}} \sum_{k,l\in\mathbb{I}} \langle\psi_i, (|\mathbf{M}_k\varphi \otimes \psi_k\rangle_{\mathbb{H}\otimes\mathbb{K}}\langle\mathbf{M}_l\varphi \otimes \psi_l|)\psi_i\rangle_{\mathbb{K}} \\
&= \sum_{i\in\mathbb{J}} |\mathbf{M}_i\varphi\rangle_{\mathbb{H}}\langle\mathbf{M}_i\varphi| = \sum_{i\in\mathbb{J}} \mathbf{M}_i\mathbf{T}\mathbf{M}_i^*.
\end{aligned}$$

This proves the result when the operator $\mathbf{T}$ is a pure state on $\mathbb{H}$. Extending this result to general trace-class operator is now easy. The last remark at the end of the theorem is now obvious for ω and is a pure state in our construction above. This proves the theorem. □

The above unitary dilation is clearly the most general possible transform for the state of a quantum system $\mathbb{H}$, such as a quantum channel Φ from $\mathbb{H}$ to $\mathbb{H}$, one could think of for a quantum system, where a quantum channel can be constructed according to Theorem 5.3.1 as follows:

- starting from an initial state $\rho \in \mathcal{S}(\mathbb{H})$, couple it to an initial state ω of any kind of environment, say system $\mathbb{K}$, that results in the coupled state $\rho \otimes \omega \in \mathcal{S}(\mathbb{H} \otimes \mathbb{K})$,
- let the coupled state $\rho \otimes \omega$ evolve in $\mathbb{H} \otimes \mathbb{K}$ driven by any unitary operator $\mathbf{U} \in \mathfrak{B}(\mathbb{H} \otimes \mathbb{K})$ that results in the quantity $\mathbf{U}(\rho \otimes \omega)\mathbf{U}^*$,
- finally, we ignore the environment $\mathbb{K}$ and look at the resulting state for $\mathbb{H}$ by partially tracing out $\mathbf{U}(\rho \otimes \omega)\mathbf{U}^*$ and obtain the expression $\mathrm{tr}_{\mathbb{K}}[\mathbf{U}(\rho \otimes \omega)\mathbf{U}^*]$.

Remark 5.1. Unitary dilations of quantum channels can also be derived from Stinespring dilation as follows. By using the Stinespring dilation (5.10), it is easy to show that any quantum channel Φ from A to itself (i. e., $B = A$) can be represented as

$$\Phi(\rho) = \mathrm{tr}_E[\mathbf{U}_\Phi(\rho \otimes \sigma_0)\mathbf{U}_\Phi^*], \tag{5.15}$$

where σ_0 is a pure state in $\mathcal{S}(\mathbb{H}_E)$ and $\mathbf{U}_\Phi$ is a unitary operator on $\mathbb{H}_{AE}$. Representation (5.15) allows one to consider any channel from a quantum system A to itself as a reduction of some unitary (reversible) evolution of the larger quantum system AE.

Remark 5.2. The unitary dilation for the general case for a given quantum channel Φ from A to B, where $B \neq A$, states (see Shirokov [156]) that one can find such quantum systems D and E' that

$$\Phi(\rho) = \mathrm{tr}_{E'}[\mathbf{U}_\Phi(\rho \otimes \sigma_0)\mathbf{U}_\Phi^*], \tag{5.16}$$

where σ_0 is a pure state in $\mathcal{S}(\mathbb{H}_D)$ and $\mathbf{U}_\Phi$ is an unitary operator from $\mathbb{H}_{AD}$ to $\mathbb{H}_{BE'}$. For an infinite-dimensional channel Φ with representation (5.10) such that $\dim(\mathbb{H}_{BE} \ominus \mathrm{range}(\mathbf{V}_\Phi)) = +\infty$, one can always take $E = E'$. This follows from the fact that any partial isometry $\mathbf{W}$ such that $\dim(\ker(\mathbf{W})) = \dim(\ker(\mathbf{W}^*)) = +\infty$ can be extended to an unitary operator (see Reed and Simon [128]).

Remark 5.3. For a given extended quantum channel $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$, the Stinespring theorem (see Theorem 4.3.1) implies existence of an environmental quantum system E represented by the Hilbert space $\mathbb{H}_E$ and of an isometry $\mathbf{V} : \mathbb{H}_A \to \mathbb{H}_B \otimes \mathbb{H}_E$ such that

$$\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}\rho\mathbf{V}^*], \quad \forall \rho \in \mathfrak{T}_+(\mathbb{H}_A). \tag{5.17}$$

For instance, taking for simplicity $A = B$, one can write

$$\Phi(\rho) = \mathrm{tr}_E[\mathbf{U}_{AE}(\rho_A \otimes \omega_E)\mathbf{U}_{AE}^*], \tag{5.18}$$

where ω_E is a fixed state of system E, U_{AE} is the unitary transformation coupling the latter to the input system A, and tr_E denotes the partial trace over the environment E. The representation (5.17) is not unique. Nonetheless, by enlarging the environment E to describe the environment state as a pure state, $\omega_E = |\omega\rangle_E\langle\omega|$ (see Section 5.5 for a proper definition of this purification mechanism), the choice of U_{AE} can be shown to be unique up to a local isometric transformation on E. Under this condition, the dilation (5.17) provides what is generally known as the Stinespring representation for Φ.

5.3.3 Kraus representation

Recall that the quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and its dual channel $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ satisfy the following *duality relation*:

$$\mathrm{tr}[\Phi(\rho_A)O_B] = \mathrm{tr}[\rho_A\Phi^*(O_B)], \quad \forall \rho_A \in \mathcal{S}(\mathbb{H}_A), O_B \in \mathfrak{B}(\mathbb{H}_B). \tag{5.19}$$

The dual channel Φ^* is linear and completely positive, but in general not trace preserving. However, it is always *unital*, i. e., it maps the identity operator $\mathbf{I}_B$ on $\mathbb{H}_B$ into the identity operator $\mathbf{I}_A$ on $\mathbb{H}_A$. In this case, operator sum representation (Kraus representation) of Φ^* can be easily derived from the Kraus representation (see equation (5.19)) of

$$\Phi(\rho_A) = \sum_j \mathbf{K}_j \rho \mathbf{K}_j^*, \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A) \tag{5.20}$$

as

$$\Phi^*(O_B) = \sum_j \mathbf{K}_j^* O_B \mathbf{K}_j, \quad O_B \in \mathfrak{B}(\mathbb{H}_B). \tag{5.21}$$

We also recall that the Stinespring representation a quantum operation $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ can be written as $\Phi(\mathbf{A}) = \operatorname{tr}_E[\mathbf{V}\mathbf{A}\mathbf{V}^*]$ for all $\mathbf{A} \in \mathfrak{T}_+(\mathbb{H}_A)$, where $\mathbf{V} : \mathbb{H}_A \to \mathbb{H}_{BE}$ is an isometry. There is a relation between the isometry $\mathbf{V}$ and operator $\mathbf{K}_i$ in the Kraus representation (5.20) as follows:

$$\langle \varphi | \mathbf{K}_i \phi \rangle_B = \langle \varphi \otimes k | \mathbf{V} \phi \rangle_{BE}, \quad \varphi \in \mathbb{H}_B, \phi \in \mathbb{H}_A,$$

where $\{|k\rangle_E\}_{k=1}^{+\infty}$ is a particular basis of $\mathbb{H}_E$.

5.4 Structural properties of quantum channels

The class of extended quantum channels $\mathfrak{EQC}(A, B)$ satisfy the following composition rules and structural properties.

5.4.1 Convexity of channels

If $\Phi, \Psi \in \mathfrak{EQC}(A, B)$ and $p \in [0, 1]$, then $p\Phi + (1 - p)\Psi \in \mathfrak{EQC}(A, B)$, where

$$\big(p\Phi + (1 - p)\Psi\big)(\rho) = p\Phi(\rho) + (1 - p)\Psi(\rho), \quad \forall \rho \in \mathfrak{T}_+(\mathbb{H}_A).$$

5.4.2 Concatenation of channels

Given $\Phi \in \mathfrak{EQC}(A, B)$ and $\Psi \in \mathfrak{EQC}(B, C)$, then their composition (or concatenation) $\Psi \circ \Phi \in \mathfrak{EQC}(A, B)$, where $\Psi \circ \Phi$ is defined by

$$\Psi \circ \Phi(\rho) = \Psi\big(\Phi(\rho)\big), \quad \forall \rho \in \mathfrak{T}_+(\mathbb{H}_A).$$

The composition $\Psi(\Phi) := \Psi \circ \Phi \in \mathfrak{CQC}(A, C)$ is called concatenation of Φ with Ψ. It physically means a serial application of channel Φ on states of system A followed by an application of the channel Ψ on the output state of Φ in system B to a state in system C.

Channels concatenation allows one to introduce a relation of equivalence between channels. In particular, two channels $\Phi, \Psi \in \mathfrak{CQC}(A)$ are said to be unitarily equivalent if there exist unitary channels $\mathfrak{U} : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_A)$ such that

$$\Psi(\rho) = (\mathfrak{U} \circ \Phi \circ \mathfrak{U}^*)(\rho), \quad \forall \rho \in \mathfrak{T}_+(\mathbb{H}_A),$$

where $\mathfrak{U}^*$ is the adjoint of $\mathfrak{U}$ and $\mathfrak{U}\mathfrak{U}^* = \mathfrak{U}^*\mathfrak{U} = \mathfrak{I}$. Note that $\mathfrak{I}$ denotes the identity operator acting on either $\mathfrak{T}_+(\mathbb{H}_A)$ or $\mathfrak{T}_+(\mathbb{H}_B)$.

5.4.3 Tensor product of channels

Given two quantum channels $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ and $\Psi : \mathfrak{T}_+(\mathbb{H}_{A'}) \to \mathfrak{T}_+(\mathbb{H}_{B'})$. Their tensor product $\Phi \otimes \Psi : \mathfrak{T}_+(\mathbb{H}_{AA'}) \to \mathfrak{T}_+(\mathbb{H}_{BB'})$ is a mapping from the composite system AA' to composite system BB' and is defined as

$$(\Phi \otimes \Psi)(\rho \otimes \omega) = \Phi(\rho) \otimes \Psi(\omega), \quad \forall \rho \in \mathfrak{T}_+(\mathbb{H}_A) \text{ and } \forall \omega \in \mathfrak{T}_+(\mathbb{H}_{A'}), \tag{5.22}$$

where $\mathbb{H}_{AA'} = \mathbb{H}_A \otimes \mathbb{H}_{A'}$, etc.

As mentioned in Caruso et al [18], concatenations and tensor products of quantum channels represent two alternative ways of composing quantum channels, which to some extent, mimic respectively the *in-series* and *in-parallel* composition rules of electrical circuit elements. In particular, channel concatenation is naturally suited to characterize the temporal correlations of a single quantum system (the sequential applications of quantum channels corresponding to different stages of the system evolution). On the contrary, the tensor product described above allows to describe spatial correlations, which might be present in the evolution of composite quantum systems. Also, tensor products can be employed to describe the transformations that a sequence of information carriers encounters when transmitted through a communication line.

5.5 Purification

Let $\rho_A \in \mathcal{S}(\mathbb{H}_A)$ be a given quantum state on $\mathbb{H}_A$ with the spectral decomposition (see Theorem 2.4.5)

$$\rho_A = \sum_{i=1}^{+\infty} \lambda_i |e_i\rangle_A \langle e_i|, \quad \lambda_i \in \mathbb{C} \text{ and } i = 1, 2, \ldots, \tag{5.23}$$

where $\{|e_i\rangle_A\}_{i=1}^{+\infty}$ is an orthonormal basis of $\mathbb{H}_A$ that consists of eigenvectors of ρ_A with corresponding eigenvalues $\{\lambda_i\}_{i=1}^{+\infty}$.

Definition 5.5.1 (Purification). A purification of ρ_A is a pure bipartite state $|\psi\rangle_{RA}$ on a reference system R and the original system A. The purification state $|\psi\rangle_{RA}$ has the property that the reduced state on system A is equal to ρ_A in equation (5.23):

$$\rho_A = \mathrm{tr}_R[|\psi\rangle_{RA}\langle\psi|], \tag{5.24}$$

where $\mathrm{tr}_R[\cdots]$ is the partial trace of $[\cdots]$ taken over the reference system R represented by the Hilbert space $\mathbb{H}_R$.

Any quantum state ρ_A has a purification $|\psi\rangle_{RA}$. The following is an example of the purification of a quantum state.

Example 5.3. Assume that the quantum state ρ_A has a spectral decomposition equation (5.23). Consider the bipartite pure state $|\psi\rangle_{RA}$ described by

$$|\psi\rangle_{RA} = \sum_{i=1}^{+\infty} \sqrt{\lambda_i}(|e_i\rangle_R \otimes |e_i\rangle_A), \tag{5.25}$$

where $(|e_i\rangle_R)_{i=1}^{+\infty}$ is an orthonormal basis for the reference system R represented by $\mathbb{H}_R$. In this case,

$$\langle\psi|_{RA} = \sum_{j=1}^{+\infty} \sqrt{\lambda_j}(\langle e_j|_R\rangle \otimes \langle e_j|_A). \tag{5.26}$$

Then $|\psi\rangle_{RA}$ is a purification of ρ_A, because

$$\begin{aligned}
&\mathrm{tr}_R[|\psi\rangle_{RA}\langle\psi|] \\
&= \mathrm{tr}_R\left[\sum_{i,j=1}^{+\infty} \sqrt{\lambda_i}\sqrt{\lambda_j}(|e_i\rangle_R \otimes |e_i\rangle_A)(\langle e_j|_A \otimes \langle e_j|_R)\right] \\
&= \mathrm{tr}_R\left[\sum_{i=1}^{+\infty} \lambda_i(|e_i\rangle_R\langle e_i|) \otimes (|e_i\rangle_A\langle e_i|)\right] = \sum_{i=1}^{+\infty} \lambda_i|e_i\rangle_A\langle e_i| = \rho_A.
\end{aligned}$$

We note that any two purifications $|\psi\rangle_{RA}$ and $|\phi\rangle_{AR}$ of ρ_A are related by some local unitary $\mathbf{U}_R$ on the reference system R by

$$|\phi\rangle_{RA} = (\mathbf{U}_R \otimes \mathbf{I}_A)|\psi\rangle_{RA}.$$

5.6 Reversible quantum channels

Following Jencova [94], Ogawa et al. [119] and Shirokov [148], we introduce the following definition of reversible quantum channels.

Definition 5.6.1. A channel $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ is *reversible* with respect to a family $\mathcal{S} \subseteq \mathcal{S}(\mathbb{H}_A)$ if there exists a channel $\Psi : \mathfrak{T}_+(\mathbb{H}_B) \to \mathfrak{T}_+(\mathbb{H}_A)$ such that $\rho = \Psi \circ \Phi(\rho)$ for all $\rho \in \mathcal{S}$. In this case, the channel Ψ will be called *a reversing channel*.

Note that the reversibility property of the channel Φ defined above is also called the *sufficiency* property of Φ (see Jencova and Petz [95] and Petz [125]).

Definition 5.6.2. A family $\mathcal{S}$ of states in $\mathcal{S}(\mathbb{H})$ is called complete if for any nonzero operator $\mathbf{A}$ in $\mathfrak{B}_+(\mathbb{H})$ there exists a state $\rho \in \mathcal{S}$ such that $\mathrm{tr}[\mathbf{A}\rho] > 0$.

Remark 5.4. A family $\{|\phi_\lambda\rangle_{\mathbb{H}}\langle\phi_\lambda|\}_{\lambda\in\Lambda}$ of pure states in $\mathcal{S}(\mathbb{H})$ is complete if and only if the linear hull of the family $\{|\phi_\lambda\rangle_{\mathbb{H}}\}_{\lambda\in\Lambda}$ is dense in $\mathbb{H}$.

We have the following result.

Lemma 5.6.3. *Let* $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ *and* $\Phi' : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_{B'})$ *be quantum channels isometrically equivalent in the sense of Definition* 5.2.2. *If the channel* Φ *is reversible with respect to a family* $\mathcal{S} \subseteq \mathcal{S}(\mathbb{H}_A)$, *then the channel* Φ' *is reversible with respect to this family* $\mathcal{S}$ *and vice versa.*

Proof. Let Ψ be a reversing channel for the channel Φ, i. e., $\Psi \circ \Phi(\rho) = \rho$ for all $\rho \in \mathcal{S}$. Consider the channel $\Theta(\cdot) = \mathbf{W}^*(\cdot)\mathbf{W} + \sigma\,\mathrm{tr}[\mathbf{I}_{B'} - \mathbf{W}\mathbf{W}^*](\cdot)$ from $\mathcal{S}(\mathbb{H}_{B'})$ into $\mathcal{S}(\mathbb{H}_B)$, where $\mathbf{W}$ is the partial isometry in (5.4), $\mathbf{I}_{B'} := \mathbf{I}_{\mathbb{H}_{B'}}$ is the identity operator on $\mathbb{H}_{B'}$ and σ is a given state in $\mathcal{S}(\mathbb{H}_B)$. It is easy to see that $\Psi \circ \Theta$ is a reversing channel for the channel Φ'. This proves the lemma. □

Definition 5.6.4. A channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is called perfectly reversible on a state $\rho \in \mathcal{S}(\mathbb{H}_A)$ if there exists a channel $\Phi : \mathcal{S}(\mathbb{H}_B) \to \mathcal{S}(\mathbb{H}_A)$ such that $\Psi \circ \Phi(\sigma) = \sigma$ for all states σ with $\mathrm{supp}(\sigma) \subseteq \mathbb{L} := \mathrm{supp}(\rho)$.

The subspace $\mathbb{L}$ mentioned in Definition 5.6.4 can be interpreted as a quantum code correcting errors of the channel Φ (see Nielsen and Chuang [116]).

Introduce the reference system $\mathbb{H}_R$ and consider a purification

$$\rho_{AR} = |\phi_{AR}\rangle_{AR}\langle\phi_{AR}| := |\phi_{AR}\rangle_{\mathbb{H}_{AR}}\langle\phi_{AR}| \in \mathcal{S}(\mathbb{H}_A \otimes \mathbb{H}_R)$$

of the state $\rho \in \mathcal{S}(\mathbb{H}_A)$.

We have the following lemma.

Lemma 5.6.5. *A channel* $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ *is perfectly reversible on a state* $\rho \in \mathcal{S}(\mathbb{H}_A)$ *if and only if there exists a channel* $\Psi : \mathcal{S}(\mathbb{H}_B) \to \mathcal{S}(\mathbb{H}_A)$ *such that*

$$((\Psi \circ \Phi) \otimes \mathfrak{I}_R)(\rho_{AR}) = \rho_{AR}. \tag{5.27}$$

Proof. Let $\Upsilon = \Psi \circ \Phi$. Consider the following set of conditions:

$$\Upsilon(|\psi\rangle_A\langle\psi|) = |\psi\rangle_A\langle\psi|, \quad \forall|\psi\rangle_A \in \mathrm{supp}(\rho) \subset \mathbb{H}_A; \tag{5.28}$$

$$\Upsilon(|\psi\rangle_A\langle\phi|) = |\psi\rangle_A\langle\phi|, \quad \forall|\psi\rangle_A, |\phi\rangle_A \in \mathrm{supp}(\rho); \subset \mathbb{H}_A \tag{5.29}$$

$$\Upsilon(|e_i\rangle_A\langle e_j|) = |e_i\rangle_A\langle e_j|, \quad \forall i, j, \tag{5.30}$$

where $\{|e_i\rangle_A\}_{i=1}^{+\infty}$ is the set of eigenvectors of the state ρ in $\mathbb{H}_A$ corresponding to nonzero eigenvalues. The following equivalences can be proved readily:

(i) Definition 5.6.4 ⇔ (5.28) follows from the spectral representation;
(ii) (5.28) ⇔ (5.29) follows from the polarization formula (1.13);
(iii) (5.29) ⇔ (5.30) is obvious;
(iv) (5.30) ⇔ (5.27) follows from the following purification formula of the state $\rho \in \mathcal{S}(\mathbb{H}_A)$:

$$|\phi_\rho\rangle_{AR} := |\phi_\rho\rangle_{\mathbb{H}_{AR}} = \sum_{i=1}^{+\infty} \sqrt{\lambda_i}(|e_i\rangle_A \otimes |e_i\rangle_R) \in \mathbb{H}_{AR}, \tag{5.31}$$

where $\mathrm{tr}_R[|\phi_\rho\rangle_{AR}] = \sum_{i=1}^{+\infty} \sqrt{\lambda_i}|e_i\rangle_A = \rho$. This proves the lemma. □

5.7 Complementary channels

The concept of complementary channels between the output and the environment of the quantum channel was first introduced by Holevo [74].

Definition 5.7.1. Let $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ be an extended quantum channel with Stinespring representation

$$\Phi(\mathbf{A}) = \mathrm{tr}_E[\mathbf{V}_\Phi \mathbf{A} \mathbf{V}_\Phi^*],$$

where $\mathbf{V}_\Phi : \mathbb{H}_A \to \mathbb{H}_B \otimes \mathbb{H}_E$ is an isometry and $\mathbb{H}_E$ is the Hilbert space representing the environment system E. Then the channel $\hat{\Phi} : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_E)$ defined by

$$\mathfrak{T}_+(\mathbb{H}_A) \ni \rho \mapsto \hat{\Phi}(\rho) = \mathrm{tr}_B[\mathbf{V}_\Phi \rho \mathbf{V}_\Phi^*] \in \mathfrak{T}_+(\mathbb{H}_E) \tag{5.32}$$

is called the *complementary channel* to the channel Φ.

The complementary channel is uniquely defined up to isometrical equivalence, i. e., if $\hat{\Phi}' : A \to E'$ is the channel defined by (5.4) via some other Stinespring isometry

$$\mathbf{V}_{\Phi'} : \mathbb{H}_A \to \mathbb{H}_B \otimes \mathbb{H}_{E'},$$

then there exists a partial isometry $\mathbf{W} : \mathbb{H}_E \to \mathbb{H}_{E'}$ such that

$$\hat{\Phi}'(\rho) = \mathbf{W}\hat{\Phi}(\rho)\mathbf{W}^* \quad \text{and} \quad \hat{\Phi}(\rho) = \mathbf{W}^*\hat{\Phi}(\rho)\mathbf{W}$$

for all $\rho \in \mathcal{S}(\mathbb{H}_A)$.

6 Approximation and convergence of quantum channels

This chapter explores different types of convergences and compares various topologies on the space of quantum channels or extended quantum channels from system A to system B. These include uniform, strong and strong* convergences as well as topologies generated by (unconstrained and energy constrained) diamond norms, (unconstrained and energy constrained) Bures distance, and completely bounded norm.

– With the exceptions of Sections 6.1, 6.5 and 6.6, it is recommended that readers skip the rest of this chapter at the first reading and revisit it when it is needed at a later time.

6.1 Distinguishability of quantum states

In the following, we study various quantities, such as fidelity and Bures distance, to measure the distinction and degree of distinction between quantum states in $\mathcal{S}(\mathbb{H})$, where $\mathbb{H}$ is a separable infinite-dimensional Hilbert space. For simplicity of exposition of these measures, we assume that the C^*-algebra $\mathcal{A}$ to be the whole space of bounded linear operators on $\mathbb{H}$, i. e., $\mathcal{A} = \mathfrak{B}(\mathbb{H})$.

6.1.1 Fidelity of quantum states

The fidelity between quantum states ρ and σ, denoted by $F(\rho,\sigma)$, is defined by

$$F(\rho,\sigma) = \operatorname{tr}[\sqrt{\rho^{1/2}\sigma\rho^{1/2}}] = \operatorname{tr}[\sqrt{\rho}\sqrt{\sigma}]. \tag{6.1}$$

We can use either one of the above two formulae as the definition of $F(\rho,\sigma)$. This is because

$$\begin{aligned}\operatorname{tr}[|\sqrt{\rho}\sqrt{\sigma}|] &= \operatorname{tr}[\sqrt{(\sqrt{\rho}\sqrt{\sigma})^*(\sqrt{\rho}\sqrt{\sigma})}] \\ &= \operatorname{tr}[\sqrt{\sqrt{\sigma}^*\sqrt{\rho}^*\sqrt{\rho}\sqrt{\sigma}}] = \operatorname{tr}[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}]\end{aligned}$$

by Proposition 1.8.4.

As shown in the following proposition, it is easy to see that $0 \le F(\rho,\sigma) \le 1$ is a measure of degree of distinguishability of the quantum states ρ from the quantum state σ, in which $F(\rho,\sigma) = 1$ means ρ is not distinct from σ and $F(\rho,\sigma) = 0$ means that ρ is totally distinct from σ, in the sense that ρ and σ are orthogonal (and denoted by $\rho \perp \sigma$ when $\rho\sigma = \sigma\rho = \rho\sigma^* = \rho^*\sigma = \mathbf{0}$).

https://doi.org/10.1515/9783110788105-006

Proposition 6.1.1. *The fidelity $F(\rho,\sigma)$ between the quantum states ρ and σ satisfies the following properties:*
1. $F(\rho,\sigma) = F(\sigma,\rho)$ *for all* $\rho,\sigma \in \mathcal{S}(\mathbb{H})$*;*
2. $0 \le F(\rho,\sigma) \le 1$ *for all* $\rho,\sigma \in \mathcal{S}(\mathbb{H})$*;*
3. $F(\rho,\sigma) = 0$ *if and only if* $\rho \perp \sigma$*;*
4. $F(\rho,\sigma) = 1$ *if and only if* $\rho = \sigma$*.*

Proof. 1. Let $\omega = \sqrt{\rho}\sqrt{\sigma}$. Since $\mathrm{tr}[\omega] = \mathrm{tr}[\omega^*]$, we have

$$\begin{aligned} F(\rho,\sigma) &= \mathrm{tr}[\sqrt{\rho}\sqrt{\sigma}] = \mathrm{tr}[(\sqrt{\rho}\sqrt{\sigma})^*] = \mathrm{tr}[(\sqrt{\sigma})^*(\sqrt{\rho})^*] \\ &= \mathrm{tr}[\sqrt{\sigma}\sqrt{\rho}] = F(\sigma,\rho). \end{aligned}$$

In the above, we have used the fact (see Theorem 1.3.3) that the square root of a positive self-adjoint operator is self-adjoint, i. e., $(\sqrt{\rho})^* = \sqrt{\rho}$ and $(\sqrt{\sigma})^* = \sqrt{\sigma}$.

2. It is obvious that $0 \le F(\rho,\sigma)$. On the other hand, we get

$$F(\rho,\sigma) = \mathrm{tr}[\sqrt{\rho}\sqrt{\sigma}] \le \frac{\mathrm{tr}[\rho] + \mathrm{tr}[\sigma]}{2} = 1,$$

where we have used Proposition 1.8.5 to obtain the inequality above.

3. Using functional calculus, we can easily prove that $\sqrt{\rho} \perp \sqrt{\sigma}$ if and only if $\rho \perp \sigma$. Hence, by the definition of fidelity, $F(\rho,\sigma) = \mathrm{tr}[\sqrt{\rho}\sqrt{\sigma}] = 0$ if and only if $\rho \perp \sigma$.

4. The sufficiency of the statement (4) is obvious from the definition of fidelity. For the converse part, let us assume $F(\rho,\sigma) = 1$. Then we obtain

$$1 = \mathrm{tr}[\sqrt{\rho}\sqrt{\sigma}] = \frac{\mathrm{tr}[\rho] + \mathrm{tr}[\sigma]}{2},$$

which is a case of equality in the arithmetic-geometric mean inequality of Proposition 1.8.5. Hence, we get $\rho = \sigma$. This proves the proposition. □

Recall that an operator $\mathbf{V} \in \mathfrak{B}(\mathbb{H})$ is called an isometry if $\mathbf{V}^*\mathbf{V} = \mathbf{I}_{\mathbb{H}}$ and $\mathbf{V}$ is called a coisometry if $\mathbf{V}\mathbf{V}^* = \mathbf{I}_{\mathbb{H}}$. If $\dim(\mathbb{H}) = +\infty$ and $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$, then by the polar decomposition (see Theorem 1.8.11), there exists an isometry or a coisometry $\mathbf{V}$ such that $\mathbf{T} = \mathbf{V}|\mathbf{T}|$, where $|\mathbf{T}| = \sqrt{\mathbf{T}^*\mathbf{T}}$. Generally speaking, $\mathbf{V}$ may not be unitary (see Definition 1.5.1 for definition of unitary operators). In fact, there exists a unitary operator $\mathbf{U}$ such that $\mathbf{T} = \mathbf{U}|\mathbf{T}|$ if and only if $\dim(\ker(\mathbf{T})) = \dim(\ker(\mathbf{T}^*))$. However, the following lemma says that this is the case if $\mathbf{T} = \mathbf{AB}$ is a product of two positive operators $\mathbf{A}$ and $\mathbf{B}$.

The presentation of the remainder of this subsection is largely based on results obtained in Hou and Qi [90].

Lemma 6.1.2. *Let* $\mathbb{H}$ *be a (separable) complex Hilbert space and* $\mathbf{A},\mathbf{B} \in \mathfrak{B}(\mathbb{H})$*. If* $\mathbf{A} \ge \mathbf{0}$ *and* $\mathbf{B} \ge \mathbf{0}$*, then there exists a unitary operator* $\mathbf{V} \in \mathfrak{B}(\mathbb{H})$ *such that* $\mathbf{AB} = \mathbf{V}|\mathbf{AB}|$*.*

Proof. To prove that $\mathbf{AB} = \mathbf{V}|\mathbf{AB}|$, we only need to show that

$$\dim(\ker(\mathbf{AB})) = \dim(\ker(\mathbf{AB})^*) = \dim(\ker(\mathbf{BA}))$$

if both $\mathbf{A}$ and $\mathbf{B}$ are positive operators. Note that, since $\mathbf{A} \geq \mathbf{0}$ and $\mathbf{B} \geq \mathbf{0}$, we have $\mathbf{AB}\varphi = 0$ if and only if $\mathbf{B}\varphi = 0$ or $\mathbf{A}\varphi = 0$ and $\mathbf{B}\varphi \neq 0$. In addition, $\ker(\mathbf{B}) \cap (\ker(\mathbf{A}) \cap \ker(\mathbf{B})^{\perp}) = \{0\}$. Therefore, we can write

$$\ker(\mathbf{AB}) = \ker(\mathbf{B}) \oplus (\ker(\mathbf{A}) \cap \ker(\mathbf{B})^{\perp}) \tag{6.2}$$

and similarly

$$\ker(\mathbf{BA}) = \ker(\mathbf{A}) \oplus (\ker(\mathbf{B}) \cap \ker(\mathbf{A})^{\perp}). \tag{6.3}$$

Obviously, if $\dim(\ker(\mathbf{A})) = \dim(\ker(\mathbf{B})) = +\infty$, then $\dim(\ker(\mathbf{AB})) = \dim(\ker(\mathbf{BA})) = +\infty$; if $\mathbf{A}$ (or $\mathbf{B}$) is injective, then

$$\dim(\ker(\mathbf{AB})) = \dim(\ker(\mathbf{BA})) = \dim(\ker(\mathbf{B}))$$

or

$$\dim(\ker(\mathbf{AB})) = \dim(\ker(\mathbf{BA})) = \dim(\ker(\mathbf{A})).$$

Assume that $\dim(\ker(\mathbf{A})) < +\infty$ and $\dim(\ker(\mathbf{B})) = +\infty$. By (6.2) and (6.3), we need only to check that $\dim(\ker(\mathbf{B}) \cap \ker(\mathbf{A})^{\perp}) = +\infty$. This is equivalent to show the following claim.

Claim: If $\mathbf{B} \geq \mathbf{0}$ and $\dim(\ker(\mathbf{B})) = +\infty$, then for any subspace $\mathbb{M} \subset \mathbb{H}$ with $\dim(\mathbb{M}) < +\infty$, $\dim(\ker(\mathbf{P}_{\mathbb{M}}\mathbf{B}\mathbf{P}_{\mathbb{M}}|_{\mathbb{M}})) = +\infty$, where $\mathbf{P}_{\mathbb{M}}$ is a projection of $\mathbb{H}$ onto $\mathbb{M}$ and $\mathbf{P}_{\mathbb{M}}\mathbf{B}\mathbf{P}_{\mathbb{M}}|_{\mathbb{M}}$ denotes the restriction of the operator $\mathbf{P}_{\mathbb{M}}\mathbf{B}\mathbf{P}_{\mathbb{M}}$ to $\mathbb{M} \subset \mathbb{H}$. In fact, by the space decomposition $\mathbb{H} = \mathbb{M} \oplus \mathbb{M}^{\perp}$, we may write

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^* & \mathbf{B}_{22} \end{pmatrix},$$

where $\mathbf{B}_{11} = \mathbf{P}_{\mathbb{M}}\mathbf{B}\mathbf{P}_{\mathbb{M}}|_{\mathbb{M}}$. Since $\mathbf{B} \geq \mathbf{0}$, there exists some contractive operator $\mathbf{D}$ such that $\mathbf{B}_{12} = \mathbf{B}_{11}^{1/2}\mathbf{D}\mathbf{B}_{22}^{1/2}$ (see, e. g., Conway [28]). Thus,

$$\ker(\mathbf{B}) = \ker(\mathbf{B}_{11}) \oplus \ker(\mathbf{B}_{22}) \oplus \mathbb{L},$$

where

$$\begin{aligned}\mathbb{L} = \{|x\rangle_{\mathbb{H}} \oplus |y\rangle_{\mathbb{H}} \mid |x\rangle_{\mathbb{H}} \in \ker(\mathbf{B}_{11})^{\perp}, |y\rangle_{\mathbb{H}} \in \ker(\mathbf{B}_{22})^{\perp}, \\ \mathbf{B}_{11}|x\rangle_{\mathbb{H}} + \mathbf{B}_{12}|y\rangle_{\mathbb{H}} = 0 \text{ and } \mathbf{B}_{12}^*|x\rangle + \mathbf{B}_{22}|y\rangle_{\mathbb{H}} = 0\}.\end{aligned}$$

Note that $\dim(\ker(\mathbf{B}_{22})) < +\infty$ and $\dim(\mathbb{L}) \leq \dim(\ker(B_{22})^{\perp}) < +\infty$, we must have $\dim(\ker(\mathbf{B}_{11})) = +\infty$. Finally, assume that both $\ker(\mathbf{A})$ and $\ker(\mathbf{B})$ are finite-dimensional. With respect to the space decomposition $\mathbb{H} = \ker(\mathbf{A})^{\perp} \oplus \ker(\mathbf{A})$, we have

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^{*} & \mathbf{B}_{22} \end{pmatrix}.$$

As $\mathbf{A}_1$ is injective with dense range,

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_1\mathbf{B}_{11} & \mathbf{A}_1\mathbf{B}_{12} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathbf{BA} = \begin{pmatrix} \mathbf{B}_{11}\mathbf{A}_1 & \mathbf{0} \\ \mathbf{B}_{12}^{*}\mathbf{A}_1 & \mathbf{0} \end{pmatrix},$$

we see that

$$\begin{aligned} &\ker(\mathbf{AB}) \\ &\quad = \{|x\rangle_{\mathbb{H}} \oplus |y\rangle_{\mathbb{H}} \mid |x\rangle_{\mathbb{H}} \in \ker(\mathbf{A})^{\perp}, |y\rangle_{\mathbb{H}} \in \ker(\mathbf{A}), \mathbf{B}_{11}|x\rangle_{\mathbb{H}} + \mathbf{B}_{12}|y\rangle_{\mathbb{H}} = 0\} \\ &\quad = \ker(\mathbf{B}_{11}) \oplus \ker(\mathbf{B}_{12}) \\ &\qquad \oplus \{|x\rangle_{\mathbb{H}} \oplus |y\rangle_{\mathbb{H}} \mid |x\rangle_{\mathbb{H}} \in \ker(\mathbf{B}_{11})^{\perp}, |y\rangle_{\mathbb{H}} \in \ker(\mathbf{B}_{12})^{\perp}, \mathbf{B}_{11}|x\rangle_{\mathbb{H}} + \mathbf{B}_{12}|y\rangle_{\mathbb{H}} = 0\} \end{aligned}$$

and

$$\ker(\mathbf{BA}) = \ker(\mathbf{A}) \oplus \{|x\rangle_{\mathbb{H}} \mid |x\rangle_{\mathbb{H}} \in (\ker(\mathbf{A})^{\perp} \cap \ker(\mathbf{B}_{11})\} = \ker(\mathbf{A}) \oplus \ker(\mathbf{B}_{11}).$$

Since

$$\begin{aligned} &\dim(\{|x\rangle_{\mathbb{H}} \oplus |y\rangle_{\mathbb{H}} \mid |x\rangle_{\mathbb{H}} \in \ker(\mathbf{B}_{11})^{\perp}, |y\rangle_{\mathbb{H}} \in \ker(\mathbf{B}_{12})^{\perp}, \mathbf{B}_{11}|x\rangle_{\mathbb{H}} + \mathbf{B}_{12}|y\rangle_{\mathbb{H}} = 0\}) \\ &\quad \leq \dim(\ker(\mathbf{B}_{12})^{\perp}), \\ &\ker(\mathbf{BA}) = \ker(\mathbf{A}) \oplus \{|x\rangle_{\mathbb{H}} \mid |x\rangle_{\mathbb{H}} \in \ker(\mathbf{A})^{\perp} \cap \ker(\mathbf{B}_{11})\} = \ker(\mathbf{A}) \oplus \ker(\mathbf{B}_{11}), \end{aligned}$$

and $\dim(\ker(\mathbf{B}_{12})) + \dim(\ker(\mathbf{B}_{12})^{\perp}) = \dim(\ker(\mathbf{A}))$, one gets $\dim(\ker(\mathbf{AB})) \leq \dim(\ker(\mathbf{BA}))$. Symmetrically, we have $\dim(\ker(\mathbf{BA})) \leq \dim(\ker(\mathbf{AB}))$ and, therefore, $\dim(\ker(\mathbf{AB})) = \dim(\ker(\mathbf{BA}))$. This completes the proof of the lemma. □

Lemma 6.1.3. *Let* $\mathbb{H}$ *be any separable complex Hilbert space and* $\mathbf{A}, \mathbf{B} \in \mathfrak{B}(\mathbb{H})$. *If* $\mathbf{A}, \mathbf{B}$ *are positive and* $\mathbf{AB} \in \mathfrak{T}(\mathbb{H})$, *then*

$$\|\mathbf{AB}\|_1 = \operatorname{tr}[|\mathbf{AB}|] = \max\{\operatorname{tr}[\mathbf{ABU}] \mid \mathbf{U} \in \mathfrak{U}(\mathbb{H})\}, \tag{6.4}$$

where $\mathfrak{U}(\mathbb{H})$ *is the unitary group of all unitary operators in* $\mathfrak{B}(\mathbb{H})$.

Proof. For any unitary operator $\mathbf{U} \in \mathfrak{U}(\mathbb{H})$, we have

$$\left|\operatorname{tr}[\mathbf{ABU}]\right| \leq \|\mathbf{U}\|_{\infty}\|\mathbf{AB}\|_1 = \|\mathbf{AB}\|_1 = \operatorname{tr}[|\mathbf{AB}|].$$

On the other hand, by Lemma 6.1.2, there exists a unitary operator $\mathbf{V}$ such that $\mathbf{AB} = \mathbf{V}|\mathbf{AB}|$. Thus,

$$|\mathbf{AB}| = \sqrt{(\mathbf{AB})^*\mathbf{AB}} = \sqrt{(\mathbf{V}^*\mathbf{AB})^2} = \mathbf{V}^*\mathbf{AB}$$

and

$$\|\mathbf{AB}\|_1 = \operatorname{tr}\big[|\mathbf{AB}|\big] = \operatorname{tr}\big[\mathbf{V}^*\mathbf{AB}\big] = \operatorname{tr}\big[\mathbf{ABV}^*\big].$$

Hence, (6.4) holds. This proves the lemma. □

Lemma 6.1.4. *Let* $\mathbb{H}$ *and* $\mathbb{K}$ *be separable infinite-dimensional complex Hilbert spaces and* $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$, $\mathbf{B} \in \mathfrak{B}(\mathbb{K})$. *Let* $(|i\rangle_{\mathbb{H}})_{i=1}^{+\infty}$, *and* $(|i\rangle_{\mathbb{K}})_{i=1}^{+\infty}$ *be any orthonormal bases of* $\mathbb{H}$ *and* $\mathbb{K}$, *respectively, and* $\mathbf{U} : \mathbb{H} \to \mathbb{K}$ *be the unitary operator defined by* $\mathbf{U}|i\rangle_{\mathbb{H}} = |i\rangle_{\mathbb{K}}$. *For each positive integer* N, *let* $|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}} = \sum_{i=1}^{N} |i\rangle_{\mathbb{H}} \otimes |i\rangle_{\mathbb{K}}$. *If* $\mathbf{A}$ *or* $\mathbf{B}$ *is a trace-class operator, then*

$$\lim_{N\to+\infty} \langle m_N|\mathbf{A}\otimes\mathbf{B}|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}} = \operatorname{tr}\big[\mathbf{UA}^*\mathbf{U}^*\mathbf{B}\big].$$

Proof. Clearly, $\mathbf{UA}^*\mathbf{U}^*\mathbf{B} \in \mathfrak{T}(\mathbb{K})$ and

$$\operatorname{tr}\big[\mathbf{UA}^*\mathbf{U}^*\mathbf{B}\big] = \sum_{i,j=1}^{+\infty} \langle i|\mathbf{UA}^*\mathbf{U}^*|j\rangle_{\mathbb{K}}\langle j|\mathbf{B}|i\rangle_{\mathbb{K}} = \sum_{i,j=1}^{+\infty} \langle i|\mathbf{A}^*|j\rangle_{\mathbb{H}}\langle j|\mathbf{B}|i\rangle_{\mathbb{K}},$$

which is absolutely convergent. Hence,

$$\lim_{N\to+\infty} \sum_{i,j=1}^{N} \langle i|\mathbf{A}^*|j\rangle_{\mathbb{H}}\langle j|\mathbf{B}|i\rangle_{\mathbb{K}} = \operatorname{tr}\big[\mathbf{UA}^*\mathbf{U}^*\mathbf{B}\big]. \tag{6.5}$$

On the other hand,

$$\begin{aligned}\langle m_N|\mathbf{A}\otimes\mathbf{B}|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}} &= \sum_{i,j=1}^{N} \big(\langle j|_{\mathbb{H}} \otimes \langle j|_{\mathbb{K}}\big)(\mathbf{A}\otimes\mathbf{B})\big(|i\rangle_{\mathbb{H}} \otimes |i\rangle_{\mathbb{K}}\big) \\ &= \sum_{i,j=1}^{N} \langle j|\mathbf{A}|i\rangle_{\mathbb{H}}\langle j|\mathbf{B}|i\rangle_{\mathbb{K}} = \sum_{i,j=1}^{N} \langle i|\mathbf{A}|j\rangle_{\mathbb{H}}\langle j||\mathbf{B}|i\rangle_{\mathbb{K}}.\end{aligned}$$

So, by (6.5), one obtains that

$$\lim_{N\to+\infty} \langle m_N|\mathbf{A}\otimes\mathbf{B}|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}} = \operatorname{tr}\big[\mathbf{UA}^*\mathbf{U}^*\mathbf{B}\big]$$

as desired. This proves the lemma. □

The following is the infinite-dimensional version of the Uhlmann's theorem (see Uhlmann [169, 170] for the original Uhlmann's theorem). Recall that a unit vector $|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}}$ is said to be a purification of a state ρ on $\mathbb{H}$ if $\rho = \mathrm{tr}_{\mathbb{K}}[|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}}\langle\psi|]$.

The infinite-dimensional version of Uhlmann's theorem was proved by Hou and Qi [90].

Theorem 6.1.5 (Infinite-dimensional Uhlmann's theorem). *Let $\mathbb{H}$ and $\mathbb{K}$ be separable infinite-dimensional complex Hilbert spaces. For any states ρ and σ on $\mathbb{H}$, we have*

$$F(\rho,\sigma) = \max\{|\langle\psi|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \mid |\psi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}_\rho, |\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}_\sigma\}, \tag{6.6}$$

where

$$\mathcal{P}_\rho = \{|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}} \mid |\psi\rangle_{\mathbb{H}\otimes\mathbb{K}} \textit{ is a purification of } \rho\}$$

and

$$\mathcal{P}_\sigma = \{|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \mid |\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \textit{ is a purification of } \sigma\}.$$

Proof. Assume that $\rho,\sigma \in \mathcal{S}(\mathbb{H})$. Then, by spectral decomposition of quantum states, there exist orthonormal bases of $\mathbb{H}$, $(|i\rangle_{\mathbb{H}})_{i=1}^{+\infty}$ and $(|j\rangle_{\mathbb{H}})_{j=1}^{+\infty}$ such that $\rho = \sum_{i=1}^{+\infty} p_i|i\rangle_{\mathbb{H}}\langle i|$ and $\sigma = \sum_{j=1}^{+\infty} q_j|j\rangle_{\mathbb{H}}\langle j|$ with $p_i > 0$, $q_j > 0$ and $\sum_{i=1}^{+\infty} p_i = \sum_{j=1}^{+\infty} q_j = 1$. If $|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}}$, $|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathbb{H}\otimes\mathbb{K}$ are purifications of ρ, σ, respectively, then there exist orthonormal sets $(|i\rangle_{\mathbb{K}})_{i=1}^{+\infty}$ and $(|j\rangle_{\mathbb{K}})_{i=1}^{+\infty}$ in $\mathbb{K}$ such that $|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}} = \sum_{i=1}^{+\infty}\sqrt{p_i}|i\rangle_{\mathbb{H}}\otimes|i\rangle_{\mathbb{K}}$ and $|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} = \sum_{j=1}^{+\infty}\sqrt{q_j}|j\rangle_{\mathbb{H}}\otimes|j\rangle_{\mathbb{K}}$. Pick any orthonormal bases $(|k\rangle_{\mathbb{H}})_{k=1}^{+\infty}$ of $\mathbb{H}$ and $(|k\rangle_{\mathbb{K}})_{k=1}^{+\infty}$ of $\mathbb{K}$. Let $\mathbf{U}_{\mathbb{H}}$, $\mathbf{U}_{\mathbb{K}}$, $\mathbf{V}_{\mathbb{H}}$ and $\mathbf{V}_{\mathbb{K}}$ be partial isometries (see Definition 5.2.2 for definition of a partial isometry) defined by

$$\mathbf{U}_{\mathbb{H}}|k\rangle_{\mathbb{H}} = |i\rangle_{\mathbb{H}}, \quad \mathbf{U}_{\mathbb{K}}|k\rangle_{\mathbb{K}} = |i\rangle_{\mathbb{K}}, \quad \mathbf{V}_{\mathbb{H}}|k\rangle_{\mathbb{H}} = |j\rangle_{\mathbb{H}}, \quad \text{and} \quad \mathbf{V}_{\mathbb{K}}|k\rangle_{\mathbb{K}} = |j\rangle_{\mathbb{K}},$$

respectively, for each $i,j,k = 1,2,\ldots$. For any integer $N > 0$, let

$$|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}} = \sum_{k=1}^{N}|k\rangle_{\mathbb{H}}\otimes|k\rangle_{\mathbb{K}}.$$

Then

$$\begin{aligned}|\psi_N\rangle_{\mathbb{H}\otimes\mathbb{K}} &:= \sum_{i=1}^{N}\sqrt{p_i}(|i\rangle_{\mathbb{H}}\otimes|i\rangle_{\mathbb{K}}) = \sum_{i=1}^{N}\sqrt{p_i}(\mathbf{U}_{\mathbb{H}}|k\rangle_{\mathbb{H}})\otimes(\mathbf{U}_{\mathbb{K}}|k\rangle_{\mathbb{K}})\\ &= \sum_{k=1}^{N}\sqrt{\rho}(\mathbf{U}_{\mathbb{H}}\otimes\mathbf{U}_{\mathbb{K}})(|k\rangle_{\mathbb{H}}\otimes|k\rangle_{\mathbb{K}}) = \sqrt{\rho}(\mathbf{U}_{\mathbb{H}}\otimes\mathbf{U}_{\mathbb{K}})|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}}\end{aligned}$$

and

$$|\phi_N\rangle_{\mathbb{H}\otimes\mathbb{K}} := \sum_{j=1}^{N} \sqrt{q_j}\big(|j\rangle_{\mathbb{H}} \otimes |j\rangle_{\mathbb{K}}\big) = \sum_{j=1}^{N} \sqrt{q_j}(\mathbf{V}_{\mathbb{H}}|k\rangle_{\mathbb{H}}) \otimes (\mathbf{V}_{\mathbb{K}}|k\rangle_{\mathbb{K}})$$
$$= \sum_{k=1}^{N} \sqrt{\sigma}(\mathbf{V}_{\mathbb{H}} \otimes \mathbf{V}_{\mathbb{K}})\big(|k\rangle_{\mathbb{H}} \otimes |k\rangle_{\mathbb{K}}\big) = \sqrt{\sigma}(\mathbf{V}_{\mathbb{H}} \otimes \mathbf{V}_{\mathbb{K}})|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}}.$$

It follows from Lemma 6.1.4 that

$$\begin{aligned}
\big|\langle\psi|\phi\rangle_{\mathbb{H}}\big| &= \lim_{N\to+\infty} \big|\langle\psi_N|\phi_N\rangle_{\mathbb{H}}\big| \\
&= \lim_{N\to+\infty} \big|\langle m_N|\mathbf{U}_{\mathbb{H}}^*\sqrt{\rho}\sqrt{\sigma}\mathbf{V}_{\mathbb{H}} \otimes \mathbf{U}_{\mathbb{K}}\mathbf{V}_{\mathbb{K}}|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}} \\
&= \big|\mathrm{tr}[\mathbf{U}\mathbf{V}_{\mathbb{H}}^*\sqrt{\rho}\sqrt{\sigma}\mathbf{U}_{\mathbb{H}}\mathbf{U}^*\mathbf{U}_{\mathbb{K}}^*\mathbf{V}_{\mathbb{K}}]\big| \le \mathrm{tr}[|\sqrt{\sigma}\sqrt{\rho}|] = \mathrm{tr}[\rho^{1/2}\sigma\rho^{1/2}] = F(\rho,\sigma),
\end{aligned} \tag{6.7}$$

where $\mathbf{U} : \mathbb{H} \to \mathbb{K}$ is the unitary operator defined by $\mathbf{U}|k\rangle_{\mathbb{H}} = |k\rangle_{\mathbb{K}}$. Therefore, we have proved that

$$F(\rho,\sigma) \ge \sup\{\big|\langle\psi|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}}\big| \mid |\psi\rangle_{\mathbb{H}\times\mathbb{K}} \in \mathcal{P}_\rho, |\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}\sigma\}.$$

Now, to complete the proof, it suffices to find $|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}_\rho$ and $|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}_\sigma$ such that $|\langle\psi|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}}| = F(\rho,\sigma)$. By applying Lemma 6.1.2, we see that $\sqrt{\sigma}\sqrt{\rho}$ has a polar decomposition $\sqrt{\sigma}\sqrt{\rho} = \mathbf{U}_0|\sqrt{\sigma}\sqrt{\rho}|$ with $\mathbf{U}_0$ a unitary operator.

Let $(|i\rangle_{\mathbb{K}})_{i=1}^{+\infty}$ be an orthonormal basis of $\mathbb{K}$ and let $|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}} = \sum_{i=1}^{+\infty} \sqrt{p_i}|i\rangle_{\mathbb{H}} \otimes |i\rangle_{\mathbb{K}}$ and $|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} = \sum_{j=1}^{+\infty} \sqrt{q_j}|j\rangle_{\mathbb{H}} \otimes |j\rangle_{\mathbb{K}}$. Then $|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}_\rho$ and $|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}_\sigma$. Let $|k\rangle_{\mathbb{H}} = |i\rangle_{\mathbb{H}}$ and $|k\rangle_{\mathbb{K}} = |i\rangle_{\mathbb{K}}$ for $i,k = 1,2,\ldots$. Then $\mathbf{U}_{\mathbb{H}} = \mathbf{I}_{\mathbb{H}}$, $\mathbf{U}_{\mathbb{K}} = \mathbf{I}_{\mathbb{K}}$, $\mathbf{V}_{\mathbb{H}}$ is a unitary operator determined by $\mathbf{V}_{\mathbb{H}}|i\rangle_{\mathbb{H}} = |j\rangle_{\mathbb{H}}$. Take $|j\rangle_{\mathbb{K}}$ so that $\mathbf{V}_{\mathbb{K}} = \mathbf{U}\mathbf{U}_0^*\mathbf{V}_{\mathbb{H}}\mathbf{U}$. Then for such choice of $|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}}$ and $|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}}$ we have

$$\begin{aligned}
\big|\langle\psi|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}}\big| &= \lim_{N\to+\infty} \big|\langle\psi_N|\phi_N\rangle_{\mathbb{H}\otimes\mathbb{K}}\big| \\
&= \lim_{N\to+\infty} \big|\langle m_N|\sqrt{\rho}\sqrt{\sigma}(\mathbf{V}_{\mathbb{H}} \otimes \mathbf{V}_{\mathbb{K}})|m_N\rangle_{\mathbb{H}\otimes\mathbb{K}}\big| = \big|\mathrm{tr}[\mathbf{U}\mathbf{V}_{\mathbb{H}}^*\sqrt{\rho}\sqrt{\sigma}\mathbf{U}^*\mathbf{V}_{\mathbb{K}}]\big| \\
&= \big|\mathrm{tr}[\mathbf{U}^*\mathbf{V}_{\mathbb{K}}\mathbf{U}\mathbf{V}_{\mathbb{H}}^*\mathbf{U}_0|\sqrt{\rho}\sqrt{\sigma}|]\big| = \big|\mathrm{tr}[|\sqrt{\rho}\sqrt{\sigma}|]\big| = F(\rho,\sigma).
\end{aligned}$$

This completes the proof of the theorem. □

By checking the proof of the above theorem, it is easily seen that the following holds.

Corollary 6.1.6. *Let $\mathbb{H}$ and $\mathbb{K}$ be separable infinite-dimensional complex Hilbert spaces. For any quantum states ρ and σ on $\mathbb{H}$, we have*

$$\begin{aligned}
F(\rho,\sigma) &= \max\{\big|\langle\psi_0|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}}\big| \mid |\phi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}_\sigma\} \\
&= \max\{\big|\langle\psi|\phi_0\rangle_{\mathbb{H}\otimes\mathbb{K}}\big| \mid |\psi\rangle_{\mathbb{H}\otimes\mathbb{K}} \in \mathcal{P}_\sigma\},
\end{aligned}$$

where $|\psi_0\rangle_{\mathbb{H}\otimes\mathbb{K}}$ *is any fixed purification of* ρ *of the form* $|\psi_0\rangle = \sum_{i=1}^{+\infty}\sqrt{p_i}|i\rangle_{\mathbb{H}}\otimes|i\rangle_{\mathbb{K}}$ *with* $\{|i\rangle_{\mathbb{K}}\}_{i=1}^{+\infty}$ *an orthonormal basis of* $\mathbb{K}$ *and* $|\phi_0\rangle_{\mathbb{H}\otimes\mathbb{K}}$ *is any fixed purification of* σ *of the form* $|\phi_0\rangle_{\mathbb{H}\otimes\mathbb{K}} = \sum_{j=1}^{+\infty}\sqrt{q_j}|j\rangle_{\mathbb{H}}\otimes|j\rangle_{\mathbb{K}}$ *with* $\{|j\rangle_{\mathbb{K}}\}_{j=1}^{+\infty}$ *being an orthonormal bases* $\mathbb{K}$.

The fidelity $F(\cdot,\cdot):\mathcal{S}(\mathbb{H})\times\mathcal{S}(\mathbb{H})\to[0,1]$ is not a distance/metric because it does not meet the triangular inequality. However, like the finite-dimensional case, by use of Theorem 6.1.5 and Corollary 6.1.6, one can show that the arc-cosine of fidelity is a distance/metric.

Corollary 6.1.7. $A(\rho,\sigma) := \arccos F(\rho,\sigma)$ *is a distance/metric on* $\mathcal{S}(\mathbb{H})$. *That is,*
(i) $0 \le A(\rho,\sigma)$ *for all* $\rho,\sigma\in\mathcal{S}(\mathbb{H})$;
(ii) $A(\rho,\sigma) = A(\sigma,\rho)$ *for all* $\sigma,\rho\in\mathcal{S}(\mathbb{H})$;
(iii) $A(\rho,\sigma) = 0$ *if and only if* $\rho=\sigma$;
(iv) $A(\rho,\omega)\le A(\rho,\sigma)+A(\sigma,\omega)$ *for all* $\rho,\sigma,\omega\in\mathcal{S}(\mathbb{H})$.

The following proposition states that the fidelity $F(\cdot,\cdot):\mathcal{S}(\mathbb{H})\times\mathcal{S}(\mathbb{H})\to[0,1]$ satisfies strong concavity condition.

Proposition 6.1.8 (Strong concavity of fidelity). *Let* $\{p_i\}_{i\in\mathbb{I}}$ *and* $\{q_i\}_{i\in\mathbb{I}}$ *be probability distributions over the same index set* $\mathbb{I}$, *and let* ρ_i *and* σ_i *be quantum states on* $\mathbb{H}$ *also indexed by the same index set. Then*

$$F\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \ge \sum_i \sqrt{p_iq_i}F(\rho_i,\sigma_i). \tag{6.8}$$

Sketch of proof. We only outline the proof, using Uhlmann's theorem (see Theorem 6.1.5), as follows. (i) Use the Uhlmann theorem to prove monotonicity of the fidelity under partial trace. (ii) Let ρ (resp., σ) denote the density operator corresponding to an ensemble of states $\{p_i,\rho_i\}_{i=1}^m$ (resp., $\{q_i,\sigma_i\}_{i=1}^m$). For each $i\in\{1,2,\dots,m\}$, let $|\psi_{\rho_i}\rangle$ (resp., $|\psi_{\sigma_i}\rangle$) denote a purification of ρ_i (resp., σ_i). (iii) Find a purification $|\psi_\rho\rangle$ of ρ (resp., $|\psi_\sigma\rangle$ of σ) in terms of the $|\psi_{\rho_i}\rangle$ (resp., $|\psi_{\sigma_i}\rangle$). (iv) Using the above, we can prove joint concavity of the fidelity, i. e.,

$$F\left(\sum_{i=1}^m p_i\rho_i, \sum_{i=1}^m q_i\sigma_i\right) \ge \sum_{i=1}^m \sqrt{p_iq_i}F(\rho_i,\sigma_i).$$

This provides an outline for the proof. Interested readers are invited to fill in the details of the proof. □

The corollary below follows from the above proposition by choosing $p_1 = p_2 = 1/2$ and $q_1 = q_2 = 1/2$.

Corollary 6.1.9 (Superadditivity). *Let* $\rho_i,\omega_i\in\mathcal{S}(\mathbb{H})$ *for* $i=1,2$. *Then*

$$F(\rho_1+\rho_2,\omega_1+\omega_2)\ge F(\rho_1,\omega_1)+F(\rho_2,\omega_2).$$

The following computational formula for infinite-dimensional fidelity will be useful later on.

Theorem 6.1.10 (Hou and Qi [90]). *Let $\mathbb{H}$ be a separable infinite-dimensional complex Hilbert space. Then, for any states $\rho, \sigma \in \mathcal{S}(\mathbb{H})$, we have*

$$F(\rho,\sigma) = \inf_{\{\mathbf{E}_m\}} F(p_m, q_m) := \inf_{\{\mathbf{E}_m\}} \sum_{m=1}^{+\infty} \sqrt{p_m}\sqrt{q_m}, \tag{6.9}$$

where the infimum is over all POVMs $\{\mathbf{E}_m\}_{m=1}^{+\infty}$ defined on $\mathbb{H}$, and $p_m = \operatorname{tr}[\rho\mathbf{E}_m]$, $q_m = \operatorname{tr}[\sigma\mathbf{E}_m]$ are the probability distributions for ρ and σ corresponding to the POVM $\{\mathbf{E}_m\}$. Moreover, the infimum attains the minimum if and only if the operator $\mathbf{M} = \rho^{[-1/2]}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\rho^{[-1/2]}$ is diagonal.

Proof. We follow the proof provided by [90] below. Let $\{\mathbf{E}_m\}_{m=1}^{+\infty}$ be a POVM defined on $\mathbb{H}$. Then $\mathbf{E}_m \geq \mathbf{0}_{\mathbb{H}}$ and $\sum_{m=1}^{+\infty}\mathbf{E}_m = \mathbf{I}_{\mathbb{H}}$; here, the series converges under the operator topology $\|\cdot\|_\infty$. By Lemma 6.1.2, there exists a unitary operator $\mathbf{U}$ such that $\sqrt{\rho^{1/2}\sigma\rho^{1/2}} = \mathbf{U}\sqrt{\sigma}\sqrt{\rho}$. Thus, by the Cauchy–Schwarz inequality and Proposition 1.8.6,

$$\begin{aligned} F(\rho,\sigma) = \operatorname{tr}[\mathbf{U}\sqrt{\sigma}\sqrt{\rho}] &= \sum_{m=1}^{+\infty}\operatorname{tr}[\mathbf{U}\sqrt{\sigma}\sqrt{\mathbf{E}_m}\sqrt{\mathbf{E}_m}\sqrt{\rho}] \\ &\leq \sum_{m=1}^{+\infty}\sqrt{\operatorname{tr}[\rho\mathbf{E}_m]\operatorname{tr}[\sigma\mathbf{E}_m]} = \sum_{m=1}^{+\infty}\sqrt{p_m q_m} = F(p_m,q_m). \end{aligned} \tag{6.10}$$

Hence, we have

$$F(\rho,\sigma) \leq \inf_{\{\mathbf{E}_m\}} F(p_m,q_m).$$

Next, we show that the equality holds in the above inequality, i. e., equation (6.9) holds. By the spectral decomposition, there is an orthonormal basis $\{|i\rangle_{\mathbb{H}}\}_{i=1}^{+\infty}$ of $\mathbb{H}$ such that $\rho = \sum_i r_i |i\rangle_{\mathbb{H}}\langle i|$ with $\sum_i r_i = 1$. For any positive integer n, let $\mathbb{H}_n$ be the n-dimensional subspace spanned by $\{|i\rangle_{\mathbb{H}}\}_{i=1}^{n}$ and $\mathbf{P}_n$ be the projection from $\mathbb{H}$ onto $\mathbb{H}_n$. Define $\rho_n = \alpha_n^{-1}\mathbf{P}_n\rho\mathbf{P}_n$ and $\sigma_n = \beta_n^{-1}\mathbf{P}_n\sigma\mathbf{P}_n$, where $\alpha_n = \operatorname{tr}[\mathbf{P}_n\rho\mathbf{P}_n]$ and $\beta_n = \operatorname{tr}[\mathbf{P}_n\sigma\mathbf{P}_n]$. Clearly, $\lim_{n\to+\infty}\alpha_n = 1$, $\lim_{n\to+\infty}\beta_n = 1$, $\lim_{n\to+\infty}\|\rho_n - \rho\|_\infty = 0$ and $\lim_{n\to+\infty}\|\sigma_n - \sigma\|_\infty = 0$.

Consequently, we see that $\lim_{n\to+\infty}\|\rho_n - \rho\|_1 = 0$ and $\lim_{n\to+\infty}\|\sigma_n - \sigma\|_1 = 0$. It follows that

$$\lim_{n\to+\infty}\left\|\sqrt{\rho_n^{1/2}\sigma_n\rho_n^{1/2}} - \sqrt{\rho^{1/2}\sigma\rho^{1/2}}\right\|_1 = 0,$$

which implies that $\lim_{n\to+\infty}F(\rho_n,\sigma_n) = F(\rho,\sigma)$. So, for any $\epsilon > 0$, there exists some N_1 such that

$$\left|F(\rho,\sigma) - \alpha_n\beta_n F(\rho_n,\sigma_n)\right| \le \epsilon/2 \tag{6.11}$$

whenever $n > N_1$.

On the other hand, note that

$$\lim_{n\to+\infty} \sqrt{\alpha_n\beta_n}\,\mathrm{tr}[\rho\mathbf{P}_n] = 1 \quad \text{and} \quad \lim_{n\to+\infty} \sqrt{\alpha_n\beta_n}\,\mathrm{tr}[\sigma\mathbf{P}_n] = 1.$$

Thus, for the above $\epsilon > 0$, there exists some N_2 such that whenever $n > N_2$.

$$\left|1 - \sqrt{\alpha_n\beta_n}\,\mathrm{tr}[\rho\mathbf{P}_n]\right| \le \epsilon/2 \quad \text{and} \quad \left|1 - \sqrt{\alpha_n\beta_n}\,\mathrm{tr}[\sigma\mathbf{P}_n]\right| \le \epsilon/2 \tag{6.12}$$

whenever $n > N_2$. Now, consider ρ_n and σ_n for $n \ge \max\{N_1, N_2\}$. With respect to the space decomposition $\mathbb{H} = \mathbb{H}_n \otimes \mathbb{H}_n^\perp$, we have

$$\rho_n = \begin{pmatrix} \rho_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \sigma_n = \begin{pmatrix} \sigma_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\rho_0, \sigma_0 \in \mathcal{S}(\mathbb{H}_n)$. Applying (6.9) to ρ_0 and σ_0, there exists POVM $\{\mathbf{E}'_m\}_{m=1}^n \subset \mathfrak{B}(\mathbb{H}_n)$ with $\sum_{m=1}^n E'_m = \mathbf{I}_n$, where $\mathbf{I}_n$ is an identity operator on $\mathbb{H}_n$ such that

$$F(\rho_0,\sigma_0) = \sum_{m=1}^n \sqrt{\mathrm{tr}[\rho_0\mathbf{E}'_m]\,\mathrm{tr}[\sigma_0\mathbf{E}'_m]}.$$

Let $\mathbf{E}_m = \mathbf{E}'_m \oplus \{0\}$ and $\mathbf{E}_{n+1} = \mathbf{I} - \mathbf{P}_n$. It is obvious that $\sum_{m=1}^{n+1} \mathbf{E}_m = \mathbf{I}$ and

$$F(\rho_n,\sigma_n) = F(\rho_0,\sigma_0) = \sum_{m=1}^n \sqrt{\mathrm{tr}[\rho_0\mathbf{E}'_m]\,\mathrm{tr}[\sigma_0\mathbf{E}'_m]} = \sum_{m=1}^{n+1} \sqrt{\mathrm{tr}[\sigma_n\mathbf{E}_m]\,\mathrm{tr}[\sigma_n\mathbf{E}_m]}.$$

Now define $\mathbf{F}_m = \sqrt{\alpha_n\beta_n}\mathbf{P}_n\mathbf{E}_m\mathbf{P}_n$ for $m = 1, 2, \cdots, n+1$ and $\mathbf{F}_0 = \mathbf{I} - \sqrt{\alpha_n\beta_n}\mathbf{P}_n$. It is clear that $\{\mathbf{F}_m\}$ is a POVM. Furthermore,

$$\begin{aligned}\sum_{m=1}^{n+1} \mathrm{tr}[\rho\mathbf{F}_m]\,\mathrm{tr}[\sigma\mathbf{F}_m] &= \sum_{m=1}^{n+1} \sqrt{\alpha_n\beta_n}\,\mathrm{tr}[\mathbf{P}_n\rho\mathbf{P}_n\mathbf{E}_m]\,\mathrm{tr}[\mathbf{P}_n\sigma\mathbf{P}_n\mathbf{E}_m] \\ &= \sum_{m=1}^{n+1} \alpha_n\beta_n \sqrt{\mathrm{tr}[\rho_n\mathbf{E}_m]\,\mathrm{tr}[\sigma_n\mathbf{E}_m]} = \alpha_n\beta_n F(\rho_n,\sigma_n)\end{aligned} \tag{6.13}$$

Hence, by (6.11)–(6.13), we get

$$\begin{aligned}&\left|F(\rho,\sigma) - \sum_{m=0}^{n+1} \sqrt{\mathrm{tr}[\rho\mathbf{F}_m]\,\mathrm{tr}[\sigma\mathbf{F}_m]}\right| \\ &\quad \le \left|F(\rho,\sigma) - \sum_{m=1}^{n+1} \sqrt{\mathrm{tr}[\rho\mathbf{F}_m]\,\mathrm{tr}[\sigma\mathbf{F}_m]}\right| + \sqrt{\mathrm{tr}[\rho\mathbf{F}_0]\,\mathrm{tr}[\sigma\mathbf{F}_0]}\end{aligned}$$

$$= \left|F(\rho,\sigma) - \alpha_n\beta_n F(\rho_n,\sigma_n)\right|$$
$$+ \sqrt{(1-\sqrt{\alpha_n\beta_n}\,\mathrm{tr}[\rho\mathbf{P}_n])(1-\sqrt{\alpha_n\beta_n}\,\mathrm{tr}[\sigma\mathbf{P}_n])} < \epsilon/2+\epsilon/2=\epsilon.$$

Thus, we have proved that, for any $\epsilon > 0$, there exists some POVM $\{\mathbf{F}_m\}$ such that

$$F(p_m, q_m) < F(\rho,\sigma) + \epsilon,$$

where $p_m = \mathrm{tr}[\rho\mathbf{F}_m]$ and $q_m = \mathrm{tr}[\sigma\mathbf{F}_m]$ are the probability distributions for ρ and σ corresponding to the POVM $\{\mathbf{F}_m\}$. So, (6.9) is true. It is clear that the infimum of (6.9) attains the minimum if and only if there exists a POVM $\{\mathbf{E}_m\}$ such that the Cauchy–Schwarz inequality is satisfied with the equality for each term in the sum of (6.10), i. e.,

$$\sqrt{\mathbf{E}_m}\sqrt{\rho} = \lambda_m\sqrt{\mathbf{E}_m}\sqrt{\rho}\mathbf{U}^* \tag{6.14}$$

for some set of numbers $\lambda_m \geq 0$. Note that

$$\sqrt{\rho^{1/2}\sigma\rho^{1/2}} = \mathbf{U}\sqrt{\sigma}\sqrt{\rho} = \sqrt{\rho}\sqrt{\sigma}\mathbf{U}^*$$

we get that the range$(\sqrt{\rho^{1/2}\sigma\rho^{1/2}}) \subset$ rangle$(\rho^{1/2})$, and hence

$$\sqrt{\sigma}\mathbf{U}^* = \rho^{[-1/2]}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}. \tag{6.15}$$

Substituting (6.15) into (6.14), we find that

$$\sqrt{\mathbf{E}_m}\sqrt{\rho} = \lambda_m\sqrt{\mathbf{E}_m}\rho^{[-1/2]}\sqrt{\rho^{1/2}\sigma\rho^{1/2}} \tag{6.16}$$

for each m. It follows that $\sqrt{\mathbf{E}_m}\sqrt{\rho} \neq \mathbf{0}$ implies that $\lambda_m \neq 0$. While, if $\sqrt{\mathbf{E}_m}\sqrt{\rho} = \mathbf{0}$, one may take $\lambda_m = 0$. Let

$$\mathbb{H}_0 = \mathrm{span}(\{\mathrm{range}(\sqrt{\mathbf{E}_m} \mid \sqrt{\mathbf{E}_m}\sqrt{\rho} = \mathbf{0})\}),$$

and $\mathbf{P}_0$ be the projection onto $\mathbb{H}_0$. Then (6.16) implies that (6.14) is equivalent to

$$\sqrt{\mathbf{E}_m}(\mathbf{I} - \mathbf{P}_0 - \lambda_m\mathbf{M}) = \mathbf{0} \tag{6.17}$$

holds for all m, where $\mathbf{M} = \rho^{[-1/2]}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\rho^{[-1/2]}$ (may be unbounded). Now it is easily seen that the closure of range$(\sqrt{\mathbf{E}_m})$ reduces $\mathbf{M}$ to the scalar operator λ_m^{-1} if $\sqrt{\mathbf{E}_m}\sqrt{\rho} \neq \mathbf{0}$, and $\ker(\mathbf{M}) = \mathbb{H}_0$. Thus, $0, \lambda_m^{-1}\sigma_p(\mathbf{M})$, the point spectrum (i. e., eigenvalues) of $\mathbf{M}$. Since $\sum_m \mathbf{E}_m = \mathbf{I}$, we see that $\bigoplus_m \mathrm{range}(\mathbf{E}_m) = \mathbb{H}$ and the spectrum of $\mathbf{M}$,

$$\sigma(\mathbf{M}) \subseteq \overline{\{0, \lambda_m^{-1}\}} = \overline{\sigma_p(\mathbf{M})}.$$

So, $\mathbf{M}$ must be diagonal. Conversely, if $\mathbf{M}$ is diagonal, say $\mathbf{M} = \sum_m \gamma_m |m\rangle_{\mathbb{H}}\langle m|$ with $\{|m\rangle_{\mathbb{H}}\}$ an orthonormal basis of $\mathbb{H}$. Let $\lambda_m = \gamma_m^{-1}$ if $\gamma_m \neq 0$; $\lambda_m = 0$ if $\gamma_m = 0$. Then the POVM $\{\mathbf{E}_m = |m\rangle_{\mathbb{H}}\langle m|\}$ satisfies (6.17), and thus (6.16). Hence,

$$F(\rho, \sigma) = \sum_m \sqrt{\operatorname{tr}[\rho \mathbf{E}_m]\operatorname{tr}[\sigma \mathbf{E}_m]} = F(p_m, q_m).$$

This completes the proof of theorem. □

The following result reveals that the trace distance and the fidelity are topologically equivalent measures of closeness for quantum states of infinite-dimensional systems.

Theorem 6.1.11. *Let* $\mathbb{H}$ *be an infinite-dimensional separable complex Hilbert space. Then*

$$1 - F(\rho, \sigma) \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)^2}, \quad \forall \rho, \sigma \in \mathcal{S}(\mathbb{H}). \tag{6.18}$$

Proof. 1. We first prove that $\frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)^2}$ for all $\rho, \sigma \in \mathcal{S}(\mathbb{H})$.

(A). Assume that both $\rho = |a\rangle_{\mathbb{H}}\langle a|$ and $\sigma = |b\rangle_{\mathbb{H}}\langle b|$ are pure states.

Then

$$\frac{1}{2}\|\rho - \sigma\|_1 = \frac{1}{2}\big\||a\rangle_{\mathbb{H}} - |b\rangle_{\mathbb{H}}\big\|_1 \leq \sqrt{1 - F\big(|a\rangle_{\mathbb{H}}, |b\rangle_{\mathbb{H}}\big)} = \sqrt{1 - F(\rho, \sigma)^2}.$$

(B). Let ρ and σ be any two states, and let $|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}}$ and $|\sigma\rangle_{\mathbb{H}\otimes\mathbb{K}}$ be purifications chosen such that $F(\rho, \sigma) = |\langle\psi|\phi\rangle_{\mathbb{H}\otimes\mathbb{K}}|$ by Theorem 6.1.5. Since the trace distance is nonincreasing under the partial trace, we see that

$$\frac{1}{2}\|\rho - \sigma\|_1 = \frac{1}{2}\||\psi\rangle_{\mathbb{H}\otimes\mathbb{K}}\|_1 \leq \sqrt{1 - F\big(|\psi\rangle_{\mathbb{H}\otimes\mathbb{K}}, |\phi\rangle_{\mathbb{H}\otimes\mathbb{K}}\big)^2} = \sqrt{1 - F(\rho, \sigma)^2}.$$

This establishes the inequality

$$\frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)^2}. \tag{6.19}$$

2. To see the other inequality (6.18) is true, Theorem 6.1.10 is needed.

For any given $\epsilon > 0$, by Theorem 6.1.10, we may take a POVM $\{\mathbf{E}_m\}$ such that

$$F(\rho, \sigma) \leq F(p_m, q_m) = \sum_m \sqrt{p_m q_m} < F(\rho, \sigma) + \epsilon, \tag{6.20}$$

where $p_m = \operatorname{tr}[\rho \mathbf{E}_m]$ and $q_m = \operatorname{tr}[\sigma \mathbf{E}_m]$ are the probabilities for obtaining outcome m for the states ρ and σ, respectively. Observe that, for both finite- and infinite-dimensional cases, we have

$$\frac{1}{2}\|\rho - \sigma\|_1 = \max_{\{\mathbf{E}_m\}} \frac{1}{2}|p_m - q_m|, \tag{6.21}$$

where the maximum is over all POVM $\{\mathbf{E}_m\}$. It follows from (6.21) and that

$$\sum_m (\sqrt{p_m} - \sqrt{q_m})^2 = \sum_m p_m + \sum_m q_m - 2F(p_m, q_m) = 2(1 - F(p_m, q_m)),$$

that

$$\begin{aligned} 2(1 - F(\rho, \sigma)) - 2\epsilon < 2(1 - F(p_m, q_m)) &= \sum_m (\sqrt{p_m} - \sqrt{q_m})^2 \\ &\leq \sum_m |\sqrt{p_m} - \sqrt{q_m}|(\sqrt{p_m} + \sqrt{q_m}) = \sum_m |p_m - q_m| \leq \|\rho - \sigma\|_1. \end{aligned}$$

Thus, we have proved that

$$(1 - F(\rho, \sigma)) - \epsilon < \frac{1}{2}\|\rho - \sigma\|_1$$

holds for any $\epsilon > 0$. This forces that

$$1 - F(\rho, \sigma) \leq \frac{1}{2}\|\rho - \sigma\|_1,$$

which, combining the inequality (6.16), completes the proof of the theorem. □

6.1.2 Bures distance

The *Bures distance* between ρ and σ, denoted by $\beta(\rho, \sigma)$, is defined as

$$\beta(\rho, \sigma) = \sqrt{2(1 - \sqrt{F(\rho, \sigma)})}, \tag{6.22}$$

where $F(\rho, \sigma)$ is the fidelity of ρ and σ defined in (6.1).

We have the following proposition, which is a direct consequence of Theorem 6.1.11.

Proposition 6.1.12. *For $\rho, \sigma \in \mathcal{S}(\mathbb{H})$, we have*

$$\frac{1}{2}\|\rho - \sigma\|_1 \leq \beta(\rho, \sigma) \leq \sqrt{\|\rho - \sigma\|_1}. \tag{6.23}$$

6.2 Channel fidelity

Based on the concepts developed in Subsection 6.1.1, we now investigate the fidelity of the channel output of the quantum states $\rho, \sigma \in \mathcal{S}(\mathbb{H}_A)$, $F(\Phi(\rho), \Phi(\sigma))$, where Φ is a quantum channel from $\mathcal{S}(\mathbb{H}_A)$ to $\mathcal{S}(\mathbb{H}_B)$. The channel output fidelity $F(\Phi(\cdot), \Phi(\cdot)) : \mathcal{S}(\mathbb{H}_A) \times \mathcal{S}(\mathbb{H}_A) \to [0, 1]$ will be referred to as channel fidelity.

6.2.1 Monotonicity of channel fidelity

Several remarkable properties of channel fidelity in the finite-dimensional case are still valid for infinite-dimensional case. For instance, the following monotonicity of the fidelity holds for both finite- and infinite-dimensional systems.

Proposition 6.2.1 (Monotonicity of channel fidelity). *Let* $\Phi \in \mathfrak{QC}(A, B)$. *Then*

$$F\big(\Phi(\rho), \Phi(\sigma)\big) \geq F(\rho, \sigma), \quad \forall \rho, \sigma \in \mathcal{S}(\mathbb{H}_A). \tag{6.24}$$

Proof. We use the Uhlmann Theorem 6.1.5 by defining a purification of the states; one can easily prove that the fidelity can be rewritten as the following semidefinite programming problem (SDP):

$$F(\rho, \sigma) = \max_{\mathbf{X}} \left\{ \frac{1}{2} \operatorname{tr}[\mathbf{X} + \mathbf{X}^*] \;\middle|\; \begin{pmatrix} \rho & \mathbf{X} \\ \mathbf{X}^* & \sigma \end{pmatrix} \geq \mathbf{0} \right\},$$

where the above maximum is taken over all $\mathbf{X} \in \mathfrak{B}(\mathbb{H}_A)$. Now let $\Phi \in \mathfrak{QC}(A, B)$ and let $\mathbf{X} \in \mathfrak{B}(\mathbb{H}_A)$ be any feasible point of the above SDP. As Φ is a complete positive map, we have

$$\begin{pmatrix} \rho & \mathbf{X} \\ \mathbf{X}^* & \sigma \end{pmatrix} \geq \mathbf{0} \Longrightarrow \begin{pmatrix} \Phi(\rho) & \Phi(\mathbf{X}) \\ \Phi(\mathbf{X}^*) & \Phi(\sigma) \end{pmatrix} \geq \mathbf{0}.$$

Furthermore, $\operatorname{tr}[\mathbf{X} + \mathbf{X}^*] = \operatorname{tr}[\Phi(\mathbf{X}) + \Phi(\mathbf{X}^*)]$ as $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is trace preserving. Thus, we can show that for every feasible point $\mathbf{X}$ of the SDP for $F(\rho, \sigma)$, we can define a feasible point $\Phi(\mathbf{X})$ of the SDP for $F(\Phi(\rho), \Phi(\sigma))$, which has the same objective value. As we are taking a maximization point over all feasible points, we have $F(\rho, \sigma) \leq F(\Phi(\rho), \Phi(\sigma))$. This proves the proposition. □

6.2.2 Ensemble average and entanglement fidelities

For the finite-dimensional case, ensemble average fidelity and entanglement fidelity are two kinds of important fidelities connected to a quantum channel. In this subsection, we give the definitions of ensemble average fidelity and entanglement fidelity connected to a quantum channel for an infinite- dimensional system, and discuss their relationship.

The presentation in this subsection is largely based on results obtained by Hou and Qi [90].

Let $\mathbb{H}$ be an infinite-dimensional separable complex Hilbert space and let a quantum channel $\Phi : \mathcal{S}(\mathbb{H}) \to \mathcal{S}(\mathbb{H})$ be a quantum channel. As in the finite-dimensional case (see, e. g., Wilde [178]), for such quantum channel Φ and a given ensemble $\{p_j, \rho_j\}_{j=1}^{+\infty}$, one can define *ensemble average fidelity* $\bar{F}(\{p_i, \rho_i\}, \Phi)$ by

$$\bar{F}(\{p_i, \rho_i\}, \Phi) = \sum_i p_i F(\rho_i, \Phi(\rho_i))^2, \tag{6.25}$$

where $F(\rho_i, \Phi(\rho_i))$ denotes the fidelity between the input state ρ_i and its channel output state $\Phi(\rho_i)$, which are both on the same Hilbert space $\mathbb{H}$.

Similarly, for a state ρ, one can define the entanglement fidelity $F_{\text{ef}}(\cdot,\cdot) : \mathcal{S}(\mathbb{H}) \times \mathfrak{QC}(\mathbb{H}) \to [0,1]$ by

$$\begin{aligned} F_{\text{ef}}(\rho, \Phi) &= F(|\psi\rangle_{\mathbb{H}\otimes\mathbb{H}}, (\Phi \otimes \mathfrak{I})(|\psi\rangle_{\mathbb{H}\otimes\mathbb{H}}\langle\psi|))^2 \\ &= \langle\psi|(\Phi \otimes \mathfrak{I})(|\psi\rangle_{\mathbb{H}\otimes\mathbb{H}}\langle\psi|)|\psi\rangle_{\mathbb{H}\otimes\mathbb{H}}, \end{aligned} \tag{6.26}$$

where $|\psi\rangle \in \mathbb{H} \otimes \mathbb{H}$ is a purification of ρ. Note that the definition of $F_{\text{ef}}(\rho, \Phi)$ does not depend on the choices of purifications. To see this, let $|\psi\rangle_{\mathbb{H}\otimes\mathbb{H}} = \sum_j \sqrt{p_j}|j\rangle_{\mathbb{H}} \otimes |\mu_j\rangle_{\mathbb{H}}$ be any purification, where $\{|j\rangle_{\mathbb{H}}\}$ is an orthonormal basis and $\{|\mu_j\rangle_{\mathbb{H}}\}$ is an orthonormal set of $\mathbb{H}$. By Kraus representation (4.12), there exists a sequence of operators $(\mathbf{E}_i)_{i=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H})$ with $\sum_i \mathbf{E}_i^* \mathbf{E}_i = \mathbf{I}_{\mathbb{H}}$ such that

$$\Phi(\sigma) = \sum_i \mathbf{E}_i \sigma \mathbf{E}_i^*, \quad \forall \sigma \in \mathcal{S}(\mathbb{H}).$$

Thus,

$$\begin{aligned} F_{\text{ef}}(\rho, \Phi) &= \sum_i \langle\phi|(\mathbf{E}_i \otimes \mathbf{I})(|\psi\rangle_{\mathbb{H}\otimes\mathbb{H}}\langle\psi|)(\mathbf{E}_i^* \otimes \mathbf{I})|\psi\rangle_{\mathbb{H}\otimes\mathbb{H}} \\ &= \sum_i \langle\phi| \sum_{j,k} \sqrt{p_j p_k}(\mathbf{E}_i \otimes \mathbf{I})(|j\rangle_{\mathbb{H}} \otimes |\mu_j\rangle_{\mathbb{H}}))\rangle_{\mathbb{H}\otimes\mathbb{H}} \\ &\quad \times (\langle k| \otimes \langle\mu_k|)(\mathbf{E}_i^* \otimes \mathbf{I})|\psi\rangle_{\mathbb{H}\otimes\mathbb{H}} \\ &= \sum_i \sum_{j,k} p_j p_k \langle j|\mathbf{E}_i|j\rangle_{\mathbb{H}} \langle k|\mathbf{E}_i^*|k\rangle_{\mathbb{H}} = \sum_i \operatorname{tr}[\mathbf{E}_i \rho], \end{aligned} \tag{6.27}$$

which is dependent only on ρ and Φ but not on its purification.

In the sequel, we will give some properties of entanglement fidelity for infinite-dimensional systems. First note that, by monotonicity of the fidelity (6.24), it is easily checked that

$$F_{\text{ef}}(\rho, \Phi) \le F^2(\rho, \Phi(\rho)), \quad \forall (\rho, \Phi) \in \mathcal{S}(\mathbb{H}) \times \mathfrak{QC}(\mathbb{H}). \tag{6.28}$$

We have the following result regarding the convexity of the function $F_{\text{ef}}(\cdot, \Phi)$.

Proposition 6.2.2. *Let* $\mathbb{H}$ *be an infinite-dimensional separable complex Hilbert space. Assume that* $\Phi : \mathcal{S}(\mathbb{H}) \to \mathcal{S}(\mathbb{H})$ *is a quantum channel. Then the entanglement fidelity* $F_{\mathrm{ef}}(\cdot, \Phi) : \mathcal{S}(\mathbb{H}) \to [0, 1]$ *is a convex function.*

Proof. Take any states $\rho_1, \rho_2 \in \mathcal{S}(\mathbb{H})$. Define a real function $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = F_{\mathrm{ef}}(x\rho_1 + (1 - x)\rho_2, \Phi), \quad \forall x \in \mathbb{R}.$$

By using (6.27) and elementary calculus, one sees that the second derivative of f is

$$f''(x) = \sum_i \left|\mathrm{tr}[(\rho_1 - \rho_2)\mathbf{E}_i]\right|^2$$

Hence, $f''(x) \geq 0$, which implies that $F(\cdot, \Phi) : \mathcal{S}(\mathbb{H}) \to [0, 1]$ is convex, as desired. □

Proposition 6.2.3. *Let* $\mathbb{H}$ *be an infinite-dimensional separable complex Hilbert space. Assume that* $\Phi : \mathcal{S}(\mathbb{H}) \to \mathcal{S}(\mathbb{H})$ *is a quantum channel. Then for any given ensemble* $\{p_j, \rho_j\}$, *we have* $F_{\mathrm{ef}}(\sum_i p_j\rho_j, \Phi) \leq \bar{F}(\{p_i, \rho_i\}, \Phi)$.

Proof. For any $k \in \mathbb{N}$, let $\lambda_k = \sum_{j=1}^{k} p_j$. Then, by Proposition 6.2.2 we have

$$\begin{aligned}
&F_{\mathrm{ef}}\left(\sum_j p_j\rho_j, \Phi\right) \\
&\quad = F_{\mathrm{ef}}\left(\lambda_k\left(\sum_{j=1}^{k} \frac{p_j}{\lambda_k}\rho_j\right) + (1 - \lambda_k)\sum_{j=k+1}^{+\infty} \frac{p_j}{1 - \lambda_k}\rho_j, \Phi\right) \\
&\quad \leq \lambda_k F_{\mathrm{ef}}\left(\sum_{j=1}^{k} \frac{p_j}{\lambda_k}\rho_j, \Phi\right) + (1 - \lambda_k)F_{\mathrm{ef}}\left(\sum_{j=k+1}^{+\infty} p_j\rho_j, \Phi\right) \\
&\quad \leq \lambda_k \sum_{j=1}^{k} \frac{p_j}{\lambda_k} F_{\mathrm{ef}}(\rho_j, \Phi) + (1 - \lambda_k)F_{\mathrm{ef}}\left(\sum_{j=k+1}^{+\infty} p_j\rho_j, \Phi\right) \\
&\quad = \sum_{j=1}^{k} p_j F_{\mathrm{ef}}(\rho_j, \Phi) + (1 - \lambda_k)F_{\mathrm{ef}}\left(\sum_{j=k+1}^{+\infty} \frac{p_j}{1 - \lambda_k}\rho_j, \Phi\right).
\end{aligned} \tag{6.29}$$

Note that $0 \leq F_{\mathrm{ef}}(\rho, \Phi) \leq 1$ and

$$\lim_{k\to+\infty} \lambda_k = \sum_{j=1}^{+\infty} p_j = 1.$$

So $\lim_{k\to+\infty}(1 - \lambda_k)F_{\mathrm{ef}}(\sum_{j=k+1}^{+\infty} \rho_j, \Phi) = 0$. Thus, for any $\epsilon > 0$, there exists some N such that whenever $k > N$. It follows from (6.29) that

$$F_{\mathrm{ef}}\left(\sum_j p_j\rho_j, \Phi\right) < \sum_{j=1}^{+\infty} p_j F_{\mathrm{ef}}(\rho_j, \Phi) + \epsilon.$$

By the arbitrariness of ϵ and (6.28), we obtain that

$$F_{\mathrm{ef}}\left(\sum_{j=1}^{+\infty} p_j\rho_j, \Phi\right) \le \sum_{j=1}^{+\infty} p_j F_{\mathrm{ef}}(\rho_j, \Phi) \le \sum_{j=1}^{+\infty} p_j F^2(\rho_j, \Phi(\rho_j)) \le \bar{F}(\{p_i, \rho_i\}, \Phi).$$

This complete the proof of the proposition. □

6.3 Norms on unconstrained channels

In this section, we explore various norms or metrics on $\mathfrak{QC}(A, B)$, the space of quantum channels from $\mathcal{S}(\mathbb{H}_A)$ to $\mathcal{S}(\mathbb{H}_B)$.

6.3.1 Diamond norm

For any quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$, we define the diamond norm $\|\Phi\|_\diamond$ of Φ as

$$\|\Phi\|_\diamond = \sup_{\rho \in \mathcal{S}(\mathbb{H}_{AR})} \left\|(\Phi \otimes \mathfrak{I}_R)(\rho)\right\|_1, \tag{6.30}$$

where R is any reference quantum system represented by the Hilbert space $\mathbb{H}_R$, $\mathbb{H}_{AR} := \mathbb{H}_A \otimes \mathbb{H}_R$, and $\mathfrak{I}_R$ is the identity operator on $\mathcal{S}(\mathbb{H}_R)$.

The diamond-norm metric between quantum channels is widely used in finite dimensions as a measure of distinguishability between these channels. But the topology (convergence) generated by the diamond-norm metric on the set of infinite-dimensional quantum channels is too strong for analysis of real variations of such channels (see Shirokov [152, 153]).

6.3.2 Complete boundedness norm

Recall from Proposition 4.1.5 that a linear map Υ from C^*-algebra $\mathcal{A}$ to another C^*-algebra $\mathcal{B}$ is completely positive if and only if $\Upsilon : \mathcal{A} \otimes \mathcal{M}_n \to \mathcal{B} \otimes \mathcal{M}_n$ is positive for each $n \in \mathbb{N}$.

The concept of complete boundedness to be defined below is closely related to completely positivity characterized in Proposition 4.1.5. The norm of complete boundedness for a linear map between two C^*-algebras $\mathcal{A}$ and $\mathcal{B}$ is defined as follows.

Definition 6.3.1. Let $\Upsilon : \mathcal{A} \to \mathcal{B}$ be a linear map between C^*-algebras $\mathcal{A}$ and $\mathcal{B}$. The norm of complete boundedness $\|\Upsilon\|_{\mathrm{cb}}$ of Υ is defined as

$$\|\Upsilon\|_{\mathrm{cb}} = \sup_{n \in \mathbb{N}} \|\Upsilon \otimes \mathbf{I}_{(n\times n)}\|_\infty, \tag{6.31}$$

where $\mathbf{I}_{(n\times n)}$ is the identity map of $n\times n$ complex matrices (i. e., $\mathbf{I}_{(n\times n)}$ is an $n\times n$ identity matrix). The map Υ is called a completely bounded map if $\|\Upsilon\|_{\text{cb}} < \infty$.

Equivalently, $\Upsilon : \mathcal{A} \to \mathcal{B}$ is completely bounded if and only if the linear map Υ_n that maps $n\times n$ matrices in $\mathcal{A}\otimes\mathcal{M}_n$ to $n\times n$ matrices in $\mathcal{B}\otimes\mathcal{M}_n$ defined by

$$\begin{pmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} & \dots & \mathbf{a}_{1n} \\ \mathbf{a}_{21} & \mathbf{a}_{22} & \dots & \mathbf{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{n1} & \mathbf{a}_{n2} & \dots & \mathbf{a}_{nn} \end{pmatrix} \mapsto \begin{pmatrix} \Upsilon(\mathbf{a}_{11}) & \Upsilon(\mathbf{a}_{12}) & \dots & \Upsilon(\mathbf{a}_{1n}) \\ \Upsilon(\mathbf{a}_{21}) & \Upsilon(\mathbf{a}_{22}) & \dots & \Upsilon(\mathbf{a}_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \Upsilon(\mathbf{a}_{n1}) & \Upsilon(\mathbf{a}_{n2}) & \dots & \Upsilon(\mathbf{a}_{nn}) \end{pmatrix} \tag{6.32}$$

is uniformly bounded for all $n\in\mathbb{N}$. That is, $\sup_{n\in\mathbb{N}}\|\Upsilon_n\|_\infty < +\infty$.

Note that any completely positive map $\Upsilon : \mathcal{A}\to\mathcal{B}$ is completely bounded, and we have

$$\|\Upsilon\|_{\text{cb}} = \|\Upsilon\|_\infty = \|\Upsilon(\mathbf{I}_{\mathcal{A}})\|_\infty = \|\mathbf{V}^*\mathbf{V}\|_\infty = \|\mathbf{V}\|_\infty^2,$$

where $\mathbf{V}$ is a Stinespring dilation for Υ, and the same operator norm notation $\|\cdot\|_\infty$ for operator on $\mathcal{A}$ and operator on $\mathbb{H}$. Obviously, $\|\Upsilon\|_\infty \le \|\Upsilon\|_{\text{cb}}$ for every completely bounded map Υ. If the range algebra $\mathcal{B}$ is Abelian, we even have equality: $\|\Upsilon\|_\infty = \|\Upsilon\|_{\text{cb}}$. If the domain algebra $\mathcal{A}$ for the positive linear map Υ is Abelian, then by Theorem 4.1.8 Υ is completely positive, but it is not sufficient to guarantee that bounded maps are completely bounded (see Paulsen [123]).

Recall from the Stinespring representation (see Theorem 4.3.1) that the linear map $\Upsilon : \mathcal{A}\to\mathcal{B}$ is completely positive if and only if it has the form

$$\Upsilon(\mathbf{a}) = \mathbf{V}^*\pi(\mathbf{a})\mathbf{V}, \quad \mathbf{a}\in\mathcal{A},$$

where $(\pi,\mathbb{K})$ is a GNS representation of $\mathfrak{B}(\mathbb{H})$ on $\mathbb{K}$ for some Hilbert space $\mathbb{K}$, and the Stinespring isometry $\mathbf{V}$ is a bounded linear operator from $\mathbb{H}$ to $\mathbb{K}$. If the map Υ is normal, then the representation π can be chosen to be normal.

The following result regarding equivalence of the operator norm of the Stinespring isometry $\mathbf{V}$ and the complete bounded norm on the completely positive map Υ is due to Kretschmann–Schlingemann–Werner [99]. The proof for this proposition is rather tedious and is omitted here. Interested readers are referred to [99] for a proof.

Proposition 6.3.2 (Kretschmann–Schlingemann–Werner [99]). *Let Υ_1 and Υ_2 be two completely positive linear maps from C^*-algebra $\mathcal{A}$ to C^*-algebra $\mathcal{B}$. Then Υ_1 and Υ_2 are closed in complete boundedness norm $\|\cdot\|_{\text{cb}}$ if and only if there exist corresponding dilations, $\mathbf{V}_1$ and $\mathbf{V}_2$, that are closed in operator norm $\|\cdot\|_\infty$. Specifically, we have*

$$\frac{\|\Upsilon_1 - \Upsilon_2\|_{\mathrm{cb}}}{\sqrt{\|\Upsilon_1\|_{\mathrm{cb}}} + \sqrt{\|\Upsilon_2\|_{\mathrm{cb}}}} \leq \inf_{\mathbf{V}_1,\mathbf{V}_2} \|\mathbf{V}_1 - \mathbf{V}_2\|_\infty \leq \sqrt{\|\Upsilon_1 - \Upsilon_2\|_{cd}}. \tag{6.33}$$

Proposition 6.3.2 generalizes the uniqueness clause in the Stinespring's dilation theorem (see Theorem 4.3.4) to complete positive maps that differ by a finite amount. As it was observed, uniqueness holds only up to partial isometries on the dilation spaces $\mathbb{H}_E$. So, one cannot expect that any two dilations satisfy such a norm bound, only that they can be chosen in a suitable way. Hence, the infimum in (6.33). The norm of complete boundedness $\|\cdot\|_{\mathrm{cb}}$ that appears in the continuity bound equation (6.33) is a stabilized version of the standard operator norm: Hence, equation (6.33) will in general fail to hold if the cb-norm $\|\cdot\|_{\mathrm{cb}}$ is replaced by the standard operator norm $\|\cdot\|_\infty$ (see [99]).

Remark 6.1. As implied in [99], the complete boundedness norm $\|\cdot\|_{\mathrm{cb}}$ is dual to the diamond-norm $\|\cdot\|_\diamond$ in the following sense $\|\Upsilon\|_{\mathrm{cb}} = \|\Upsilon^*\|_\diamond$, where $\Upsilon^* : \mathcal{B}^* \to \mathcal{A}^*$ denotes the dual map of Υ. This duality is frequently used in the realm of quantum computing (see Aharonov–Kitaev–Nisan [1]). Obviously, by replacing the maps in equation (6.33) by its dual counterparts, the analogous bounds hold for the diamond norm as well. The continuity bound equation (6.33) shows that the distance between two completely positive maps can equivalently be evaluated in terms of their dilations. We call this distance measure the Bures distance, which is the subject of discussion in the following subsection.

6.3.3 Bures distance

The Bures distance evaluates the distance between two complete positive maps, quantum channels and dual quantum channels, in particular, in terms of their dilations. We first discuss dual quantum channel $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$.

Recall from (6.22) that the Bures distance between quantum states ρ and σ is defined as

$$\beta(\rho,\sigma) = \sqrt{2(1 - \sqrt{F(\rho,\sigma)})},$$

where $F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1$ is the *fidelity* of ρ and σ.

Let $\Phi, \Psi \in \mathfrak{QC}(A,B)$. We define the Bures distance between quantum channels Φ and Ψ as

$$\beta(\Phi,\Psi) := \sup \beta\big((\Phi \otimes \mathfrak{I}_R)(\rho), (\Psi \otimes \mathfrak{I}_R)(\rho)\big), \quad \forall \rho \in \mathcal{S}(\mathbb{H}_{AR}) \tag{6.34}$$

between quantum channels Φ and Ψ, where $\beta(\cdot,\cdot)$ in the right-hand side is the Bures distance between quantum states defined in equation (6.22) and R is any system.

Proposition 6.3.3. *For any channels Φ and Ψ, we have*

$$\frac{1}{2}\|\Phi - \Psi\|_\diamond \le \beta(\Phi, \Psi) \le \sqrt{\|\Phi - \Psi\|_\diamond}. \tag{6.35}$$

Proof. This follows immediately from the definition of diamond-norm $\|\cdot\|_\diamond$ (6.30), the definition of Bures distance $\beta(\cdot,\cdot)$ (6.34) and inequality (6.22),

$$\frac{1}{2}\|\rho - \sigma\|_1 \le \beta(\rho, \sigma) \le \sqrt{\|\rho - \sigma\|_1}.$$

This proves the proposition. □

Inequality (6.35) indicates that the topology generated by the diamond norm is topologically equivalent to the Bures distance.

Remark 6.2. As mentioned in Shirokov [152] and [153], the topology generated by the diamond norm on the set of infinite-dimensional quantum channels is too strong for analysis of real variations of such channels: there are infinite-dimensional channels with close physical parameters such that the diamond-norm distance between them equals to 2. In this case, it is natural to use the substantially weaker topology of strong convergence on the set of quantum channels defined by the family of seminorms $\Phi \mapsto \|\Phi(\rho)\|_1$, $\rho \in \mathcal{S}(\mathbb{H}_A)$. The convergence of a sequence $(\Phi_n)_{n=1}^{+\infty}$ of channels to a channel Φ_0 in this topology means that

$$\lim_{n\to+\infty} \Phi_n(\rho) = \Phi_0(\rho), \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A).$$

An equivalent definition of the Bures distance between two quantum channels $\Phi, \Psi \in \mathfrak{QC}(A, B)$ is given below.

Definition 6.3.4. The Bures distance $\beta(\Phi, \Psi)$ between channels Φ and Ψ can be written as

$$\beta(\Phi, \Psi) = \inf \|\mathbf{V}_\Phi - \mathbf{V}_\Psi\|_\infty \tag{6.36}$$

where the infimum is taken over all common Stinespring representations

$$\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}_\Phi \rho \mathbf{V}_\Phi^*] \quad \text{and} \quad \Psi(\rho) = \mathrm{tr}_E[\mathbf{V}_\Psi \rho \mathbf{V}_\Psi^*], \tag{6.37}$$

where $\mathbf{V}_\Phi$ and $\mathbf{V}_\Psi$ are isometries from $\mathbb{H}_A$ to $\mathbb{H}_{BE}$ and $\mathbf{V}_\Phi^*$ and $\mathbf{V}_\Psi^*$ are their adjoints.

Proposition 6.3.5. *The Bures distance $\beta(\cdot,\cdot) : \mathfrak{QC}(A, B) \times \mathfrak{QC}(A, B) \to \mathbb{R}_+$ defined in* (6.36) *is a metric.*

Proof. Positivity and symmetry are immediate from (6.36). Obviously, $\beta(\Phi, \Phi) \ge 0$. Conversely, Proposition 6.3.2 shows that $\beta(\Phi, \Psi) = 0$ entails $\|\Phi - \Psi\|_{\mathrm{cb}} = 0$, and hence, $\Phi = \Psi$. Thus, it only remains to establish the triangle inequality,

$$\beta(\Phi, \Upsilon) \leq \beta(\Phi, \Psi) + \beta(\Psi, \Upsilon), \quad \forall \Phi, \Psi, \Upsilon \in \mathfrak{QC}(A, B).$$

Let $\mathbf{V}_\Phi$, $\mathbf{V}_\Psi$ and $\mathbf{V}_\Upsilon : \mathbb{H}_A \to \mathbb{H}_{BE}$ be isometries of the Stinespring representations of Φ, Ψ and Υ, respectively. That is, for all $\rho \in \mathcal{S}(\mathbb{H}_A)$,

$$\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}_\Phi \rho \mathbf{V}_\Phi^*], \quad \Psi(\rho) = \mathrm{tr}_E[\mathbf{V}_\Psi \rho \mathbf{V}_\Psi^*] \quad \text{and} \quad \Upsilon(\rho) = \mathrm{tr}_E[\mathbf{V}_\Upsilon \rho \mathbf{V}_\Upsilon^*].$$

Then

$$\begin{aligned} \beta(\Phi, \Upsilon) &\leq \|\mathbf{V}_\Phi - \mathbf{V}_\Upsilon\|_\infty \\ &\leq \|\mathbf{V}_\Phi - \mathbf{V}_\Psi\|_\infty + \|\mathbf{V}_\Psi - \mathbf{V}_\Upsilon\|_\infty \\ &= \|\tilde{\mathbf{V}}_\Phi - \tilde{\mathbf{V}}_\Psi\|_\infty + \|\tilde{\mathbf{V}}_\Psi - \tilde{\mathbf{V}}_\Upsilon\|_\infty \\ &= \beta(\Phi, \Psi) + \beta(\Psi, \Upsilon), \end{aligned}$$

where $\tilde{\mathbf{V}}_\Phi$, $\tilde{\mathbf{V}}_\Psi$ and $\tilde{\mathbf{V}}_\Upsilon$ are the corresponding minimal Stinespring dilations. This proves the triangle inequality. Therefore, $\beta(\cdot, \cdot)$ is a metric on $\mathfrak{QC}(A, B)$. □

6.4 Norms on constrained channels

Due to infinite dimensionality of quantum systems whose Hamiltonian takes the form of an $\mathfrak{H}$-operator, situations often happen that the classical and quantum capacity (to be discussed in later chapters) become infinite when quantum channels are repeatedly used. To prevent this from happening, we consider energy constrained quantum channels that apply to a compact subset $\mathcal{K}_\mathbf{H}(E)$ of $\mathcal{S}(\mathbb{H})$.

In the following, let $\mathbf{H}$ be an $\mathfrak{H}$-operator on the input space $\mathbb{H}_A$ (see (3.2) for the definition where λ_n, $n \geq 0$ were used for the point spectrum of $\mathbf{H}$ instead of E_n used here for representing energy level $\mathbf{H}$) that represent a Hamiltonian of the quantum system A. That is, $\mathbf{H}$ is an unbounded positive linear operator on $\mathbb{H}_A$ with discrete spectrum (eigenvalues) of finite multiplicity:

$$0 \leq E_0 \leq E_1 \leq \cdots \leq E_n \leq \cdots$$

with the smallest eigenvalue denoted by $E_0 = \inf_{\|\phi\|_{\mathbb{H}_A}=1} \langle \phi | \mathbf{H} | \phi \rangle_{\mathbb{H}_A}$.

For each $E > E_0$, let $\mathcal{K}_\mathbf{H}(E)$ be the compact subset of $\mathcal{S}(\mathbb{H}_A)$ defined by

$$\mathcal{K}_\mathbf{H}(E) = \{\rho \in \mathcal{S}(\mathbb{H}_A) \mid \mathrm{tr}[\mathbf{H}\rho] \leq E\}. \tag{6.38}$$

In this section, various energy-constrained norms/distances, including operator norm, diamond norm and Bures distance, for quantum channels are explored.

6.4.1 Energy-constrained operator norm

Recall that the operator norm $\|\mathbf{A}\|_\infty$ of $\mathbf{A} \in \mathfrak{B}(\mathbb{H}_A)$ is defined as

$$\|\mathbf{A}\|_\infty = \sup_{\varphi \in \mathbb{H}_A} \frac{\|\mathbf{A}\varphi\|_{\mathbb{H}_A}}{\|\varphi\|_{\mathbb{H}_A}} = \sup_{\|\varphi\|_{\mathbb{H}_A}=1} \|\mathbf{A}\varphi\|_{\mathbb{H}_A}.$$

In this section, we define an energy-constrained operator norm, denoted by $\|\mathbf{A}\|_E$, based on the energy level of an $\mathfrak{H}$-operator $\mathbf{H}$. The topology generated by $\|\cdot\|_E$ on $\mathfrak{B}(\mathbb{H}_A)$ will be compared to that generated by the operator norm $\|\cdot\|_\infty$.

For given $E > E_0$, consider the function $\|\cdot\|_E : \mathfrak{B}(\mathbb{H}_A) \to \mathbb{R}_+$ defined by

$$\|\mathbf{A}\|_E = \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} \sqrt{\operatorname{tr}[\mathbf{A}\rho\mathbf{A}^*]}, \quad \mathbf{A} \in \mathfrak{B}(\mathbb{H}_A), \tag{6.39}$$

Proposition 6.4.1. *The function $\|\cdot\|_E : \mathfrak{B}(\mathbb{H}_A) \to \mathbb{R}_+$ defined in* (6.39) *is a norm.*

Proof. To show that $\|\cdot\|_E : \mathfrak{B}(\mathbb{H}_A) \to \mathbb{R}_+$ is a norm, we need to prove that (i) $\|c\mathbf{A}|_E = |c|\|\mathbf{A}\|_E$ for all $c \in \mathbb{C}$ and $\mathbf{A} \in \mathfrak{B}(\mathbb{H}_A)$; (ii) $\|\mathbf{A} + \mathbf{B}\|_E \le \|\mathbf{A}\|_E + \|\mathbf{B}\|_E$ for all $\mathbf{A}, \mathbf{B} \in \mathfrak{B}(\mathbb{H}_A)$; and (iii) $\|\mathbf{A}\|_E \ge 0$ for all $\mathbf{A} \in \mathfrak{B}(\mathbb{H}_A)$ and $\|\mathbf{A}\|_E = 0$ implies $\mathbf{A} = \mathbf{0}$. The fulfillment of these three conditions are given below.

(i) It is clear that for all $c \in \mathbb{C}$ and $\mathbf{A} \in \mathfrak{B}(\mathbb{H}_A)$,

$$\begin{aligned}\|c\mathbf{A}\|_E &= \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} \sqrt{\operatorname{tr}[c\mathbf{A}\rho(c\mathbf{A})^*]} = \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} \sqrt{\operatorname{tr}[\bar{c}c\mathbf{A}\rho\mathbf{A}^*]} \\ &= |c| \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} \sqrt{\operatorname{tr}[\mathbf{A}\rho\mathbf{A}^*]} = |c|\|\mathbf{A}\|_E.\end{aligned}$$

(ii) Let $\mathbf{A}, \mathbf{B} \in \mathfrak{B}(\mathbb{H}_A)$, we want to show the triangle inequality

$$\|\mathbf{A} + \mathbf{B}\|_E \le \|\mathbf{A}\|_E + \|\mathbf{B}\|_E$$

holds. By the definition of $\|\mathbf{A} + \mathbf{B}\|_E$, we take for any $\epsilon > 0$ there exist a state $\rho \in \mathcal{K}_{\mathbf{H}}(E)$ such that

$$\begin{aligned}\|\mathbf{A} + \mathbf{B}\|_E &\le \sqrt{\operatorname{tr}[(\mathbf{A} + \mathbf{B})\rho(\mathbf{A} + \mathbf{B})^*]} + \epsilon \\ &= \sqrt{\operatorname{tr}[(\mathbf{A} + \mathbf{B})^*(\mathbf{A} + \mathbf{B})\rho]} + \epsilon \quad \text{(by Proposition 1.8.4)} \\ &\le \sqrt{\operatorname{tr}[|\mathbf{A} + \mathbf{B}|^2\rho]} + \epsilon.\end{aligned}$$

Then, by using the spectral decomposition of ρ, basic properties of the norm in $\mathbb{H}_A$ and the Cauchy–Schwarz inequality (1.2), it is easy to show that

$$\sqrt{\operatorname{tr}[|\mathbf{A} + \mathbf{B}|^2\rho]} \le \sqrt{\operatorname{tr}[|\mathbf{A}|^2\rho]} + \sqrt{\operatorname{tr}[|\mathbf{B}|^2\rho]} \le \|\mathbf{A}\|_E + \|\mathbf{B}\|_E.$$

Therefore, we have for any arbitrary small $\epsilon \geq 0$,

$$\|\mathbf{A}+\mathbf{B}\|_E \leq \|\mathbf{A}\|_E + \|\mathbf{B}\|_E + \epsilon.$$

This proves the triangle inequality $\|\mathbf{A}+\mathbf{B}\|_E \leq \|\mathbf{A}\|_E + \|\mathbf{B}\|_E$.

(iii) It is clear that $\|\mathbf{A}\|_E \geq 0$ for all $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$ and $\|\mathbf{A}\|_E = 0$ implies that $\mathbf{A} = \mathbf{0}$ by the definition of $\|\cdot\|_E$. This proves the proposition. □

Proposition 6.4.2. *For any given $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$, the following statements hold:*
1. $\lim_{E\to+\infty} \|\mathbf{A}\|_E = \|\mathbf{A}\|_\infty$.
2. *the map $E \mapsto \|\mathbf{A}\|_E$ is concave and nondecreasing on $[E_0, +\infty[$.*
3. $\|\mathbf{A}\varphi\|_{\mathbb{H}} \leq K_\varphi \|\varphi\|_{\mathbb{H}}$ *for all unit vectors $\varphi \in \mathbb{H}$ with finite $E_\varphi := \langle\varphi|\mathbf{A}|\varphi\rangle_{\mathbb{H}}$, where $K_\varphi = 1$ if $E_\varphi \leq E$ and $K_\varphi = \sqrt{(E_\varphi - E_0)\backslash(E-E_0)}$, otherwise.*

Proof. Let $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$. Then:

$$1. \quad \lim_{E\to+\infty} \|\mathbf{A}\|_E = \lim_{E\to+\infty} \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} \sqrt{\operatorname{tr}[\mathbf{A}\rho\mathbf{A}^*]}$$
$$= \sup_{\rho\in\mathcal{S}(\mathbb{H})} \sqrt{\operatorname{tr}[\mathbf{A}\rho\mathbf{A}^*]} \quad \Big(\text{because } \lim_{E\to+\infty} \mathcal{K}_{\mathbf{H}}(E) = \mathcal{S}(\mathbb{H})\Big)$$
$$= \sup_{\rho\in\mathcal{S}(\mathbb{H})} \frac{\|\mathbf{A}\rho\|_1}{\|\rho\|_1} = \|\mathbf{A}\|_\infty.$$

2. It is clear that the map $E \mapsto \|\mathbf{A}\|_E$ is nondecreasing on $[E_0, +\infty[$ by its definition. To show the concavity of the map $E \mapsto \|\mathbf{A}\|_E$, let E_1 and E_2 be such that $E_0 \leq E_1 < E_2 < \infty$ and $0 \leq \lambda \leq 1$, and we have

$$\|\mathbf{A}\|_{\lambda E_1+(1-\lambda)E_2} = \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(\lambda E_1+(1-\lambda)E_2)} \sqrt{\operatorname{tr}[\mathbf{A}\rho\mathbf{A}^*]} \leq \lambda\|\mathbf{A}\|_{E_1} + (1-\lambda)\|\mathbf{A}\|_{E_2}.$$

This shows that the map $E \mapsto \|\mathbf{A}\|_E$ is concave.

3. Let φ be any unit vector in $\mathbb{H}$ with finite $E_\varphi := \langle\varphi|\mathbf{H}|\varphi\rangle_{\mathbb{H}}$. Let $\epsilon > 0$ be an arbitrarily small positive number and

$$\rho = (1-K_\varphi^{-2})|\varphi_\epsilon\rangle_{\mathbb{H}}\langle\varphi_\epsilon| + K_\varphi^{-2}|\varphi\rangle_{\mathbb{H}}\langle\varphi|,$$

where φ_ϵ is a vector in $\mathbb{H}$ such that $\langle\varphi_\epsilon|\mathbf{H}|\varphi_\epsilon\rangle_{\mathbb{H}} \leq E_0 + \epsilon$. We have

$$\operatorname{tr}[\mathbf{H}\rho] = (1-K_\varphi^{-2})\operatorname{tr}[\mathbf{H}|\varphi_\epsilon\rangle_{\mathbb{H}}\langle\varphi_\epsilon|] + K_\varphi^{-2}\operatorname{tr}[\mathbf{H}|\varphi\rangle_{\mathbb{H}}\langle\varphi|]$$
$$= (1-K_\varphi^{-2})\operatorname{tr}[\langle\varphi_\epsilon|\mathbf{H}|\varphi_\epsilon\rangle_{\mathbb{H}}] + K_\varphi^{-2}\operatorname{tr}[\langle\varphi|\mathbf{H}|\varphi\rangle_{\mathbb{H}}]$$
$$\leq (1-K_\varphi^{-2})(E_0+\epsilon) + K_\varphi^{-2}E_\varphi \leq E + \epsilon,$$

and hence,

$$K_\varphi^{-1}\|\mathbf{A}\varphi\|_{\mathbb{H}} \leq \sqrt{\mathbf{A}\rho\mathbf{A}^*} \leq \|\mathbf{A}\|_{E+\epsilon}.$$

By passing to the limit $\epsilon \downarrow 0$, we obtain the required inequality. This proves the proposition. □

6.4.2 Constrained diamond norms and Bures distances

Let $\mathbf{H}$ be any $\mathfrak{H}$-operator. The energy-constrained diamond norm of quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is defined by

$$\|\Phi\|_\diamond^E = \sup_{\rho\in\mathcal{S}(\mathbb{H}_{AR}),\mathrm{tr}[\mathbf{H}\rho_A]\leq E} \left\|(\Phi\otimes\mathfrak{I}_R)\rho\right\|_1, \quad E > E_0, \tag{6.40}$$

where R is any quantum system and $\rho_A = \mathrm{tr}_R[\rho_{AR}]$.

We define the *energy-constrained Bures distance* between the quantum channels Φ and $\Psi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ as

$$\beta_E(\Phi,\Psi) = \sup_{\rho\in\mathcal{S}(\mathbb{H}_{AR}),\mathrm{tr}[\mathbf{H}\rho_A]\leq E} \beta\big((\Phi\otimes\mathfrak{I}_R)\rho, (\Psi\otimes\mathfrak{I}_R)\rho\big), \quad E > E_0, \tag{6.41}$$

where $\beta(\cdot,\cdot)$ is the Bures distance defined in equation (6.34).

Proposition 6.4.3. *The following inequalities hold:*

$$\frac{1}{2}\|\Phi-\Psi\|_\diamond^E \leq \beta_E(\Phi,\Psi) \leq \sqrt{\|\Phi-\Psi\|_\diamond^E} \tag{6.42}$$

Proof. The inequality follows easily from the definition energy-constrained diamond norm $\|\cdot\|_\diamond^E$ (see equation (6.40)), the definition of energy-constrained Bures distance $\beta_E(\cdot,\cdot)$ (see equation (6.41)) and inequality (6.35). This proves the proposition. □

Lemma 6.4.4. *Let $\mathbf{H}$ be an $\mathfrak{H}$-operator on $\mathbb{H}_A$ and let $E > E_0$, where $E_0 := \inf_{\|\varphi\|_{\mathbb{H}_A}=1}\langle\varphi|\mathbf{H}|\varphi\rangle_A$ is the smallest eigenvalue of $\mathbf{H}$. For any arbitrary quantum channel Φ from A to B, there exist a separable Hilbert space $\mathbb{H}_E$ and a Stinespring isometry $\mathbf{V}_\Phi : \mathbb{H}_A \to \mathbb{H}_{BE}$ of the channel Φ with the following property: for any quantum channel Ψ from A to B, there is a Stinespring isometry $\mathbf{V}_\Psi : \mathbb{H}_A \to \mathbb{H}_{BE}$ of Ψ such that*

$$\|\mathbf{V}_\Psi - \mathbf{V}_\Phi\|_E = \beta_E(\Psi,\Phi).$$

Proof. A sketch of the proof Let $\mathbf{V}_\Phi$ be the isometry from any Stinespring representation with infinite-dimensional environment space $\mathbb{H}_E$ and $\tilde{\mathbf{V}}_\Phi$ the isometry from $\mathbb{H}_A$ into

$$\mathbb{H}_B\otimes(\mathbb{H}_{E1}\oplus\mathbb{H}_{E2}) = (\mathbb{H}_B\otimes\mathbb{H}_{E1})\oplus(\mathbb{H}_B\otimes\mathbb{H}_{E2}),$$

where $\mathbb{H}_{E1}$ and $\mathbb{H}_{E2}$ are copies of $\mathbb{H}_E$, defined by setting $\tilde{\mathbf{V}}_\Phi|\varphi\rangle_A = \mathbf{V}_\Phi|\varphi\rangle_A \oplus |0\rangle_E$ for any $\varphi \in \mathbb{H}_A$. Since any separable Hilbert space can be isometrically embedded into $\mathbb{H}_E$, we may assume that any channel Ψ from A to B has a Stinespring representation with the same environment space $\mathbb{H}_E$. Denote by $\mathbf{V}_\Psi$ the Stinespring isometry of the channel Ψ in this representation. The arguments from the proof of Proposition 1 in Shirokov [154] (obtained by simple modification of the proof of Theorem 1 in Kretschmann–Schlingemann–Werner [99]) show that $\beta_E(\Psi, \Phi) = \|\mathbf{V}_\Psi - \mathbf{V}_\Phi\|_E$ for the Stinepring isometry $\tilde{\mathbf{V}}_\Psi : \mathbb{H}_A \to \mathbb{H}_B \otimes (\mathbb{H}_{E1} \oplus \mathbb{H}_{E2})$ of the channel Ψ defined by setting

$$\tilde{\mathbf{V}}_\Psi|\varphi\rangle = (\mathbf{I}_B \otimes \mathbf{C}_\Psi)\mathbf{V}_\Psi|\varphi\rangle \oplus \left(\mathbf{I}_B \otimes \sqrt{\mathbf{I}_E - \mathbf{C}_\Psi^*\mathbf{C}_\Psi}\right)\mathbf{V}_\Psi|\varphi\rangle$$

for any $|\varphi\rangle \in \mathbb{H}_A$, where $\mathbf{C}_\Psi \in \mathfrak{B}(\mathbb{H}_E)$ is a particular contraction. This implies the assertion of the lemma with the isometry $\tilde{\mathbf{V}}_\Psi$ in the role of $\mathbf{V}_\Psi$. This provides a sketch of the proof. □

Proposition 6.4.5. *Let* $\mathbf{H}$ *be any* $\mathfrak{H}$*-operator defined on* $\mathbb{H}_A$. *For any* $E > E_0$, *the function* $(\Phi, \Psi) \mapsto \beta_E(\Phi, \Psi)$ *defined in* (6.41) *is a real metric on* $\mathfrak{QC}(A, B)$, *which can be represented as follows:*

$$\beta_E(\Phi, \Psi) = \inf\Big(\sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} \sqrt{\operatorname{tr}[(\mathbf{V}_\Phi - \mathbf{V}_\Psi)\rho(\mathbf{V}_\Phi^* - \mathbf{V}_\Psi^*)]}\Big), \tag{6.43}$$

where the infimum is over all common Stinespring representations

$$\Phi(\rho) = \operatorname{tr}_E[\mathbf{V}_\Phi\rho\mathbf{V}_\Phi^*] \quad \textit{and} \quad \Psi(\rho) = \operatorname{tr}_E[\mathbf{V}_\Psi\rho\mathbf{V}_\Psi^*]. \tag{6.44}$$

Proof. We want to show the right-hand side of (6.43) equals the right-hand side of (6.41). Denote by $\beta'_E(\Phi, \Psi)$ the right-hand side of (6.43). Let $\mathcal{K}_{\mathbf{H}}(E)$ be the subset of $\mathcal{S}(\mathbb{H}_A)$ defined by (3.2) and let $\mathcal{N}(\Phi, \Psi) = \bigcup \mathbf{V}_\Phi^*\mathbf{V}_\Psi$, where the union is over all common Stinespring representations (6.44). Then it is easy to see that

$$\begin{aligned}\beta'_E(\Phi, \Psi) &= \inf\Big(\sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} \sqrt{\operatorname{tr}[(\mathbf{V}_\Phi - \mathbf{V}_\Psi)\rho(\mathbf{V}_\Phi^* - \mathbf{V}_\Psi^*)]}\Big) \\ &= \inf_{\mathbf{N}\in\mathcal{N}(\Phi,\Psi)} \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} \sqrt{2 - 2\Re(\operatorname{tr}[\mathbf{N}\rho])}.\end{aligned} \tag{6.45}$$

In the following, we show that $\mathcal{N}(\Phi, \Psi) = \mathcal{M}(\Phi, \Psi)$, where

$$\mathcal{M}(\Phi, \Psi) = \{\mathbf{V}_\Phi^*(\mathbf{I}_B \otimes \mathbf{C})\mathbf{V}_\Psi \mid \mathbf{C} \in \mathfrak{B}(\mathbb{H}_E), \|\mathbf{C}\|_\infty \le 1\},$$

defined via some fixed common Stinespring representation (6.44). It would imply, in particular, that $\mathcal{M}(\Phi, \Psi)$ does not depend on this representation.

To show that $\mathcal{M}(\Phi, \Psi) \subseteq \mathcal{N}(\Phi, \Psi)$, it suffices to find for any contraction $\mathbf{C} \in \mathfrak{B}(\mathbb{H}_E)$ a common Stinespring representation for Φ and Ψ with the isometries $\tilde{\mathbf{V}}_\Phi$ and $\tilde{\mathbf{V}}_\Psi$ from

$\mathbb{H}_A$ to $\mathbb{H}_B \otimes \mathbb{H}_{\tilde{E}}$ such that $\tilde{\mathbf{V}}_\Phi^* \tilde{\mathbf{V}}_\Psi = \tilde{\mathbf{V}}_\Phi(\mathbf{I}_B \otimes \mathbf{C})\mathbf{V}_\Psi$. Let $\mathbf{H}_{\tilde{E}} = \mathbb{H}_E^1 \otimes \mathbb{H}_E^2$, where $\mathbb{H}_E^1$ and $\mathbb{H}_E^2$ are copies of $\mathbb{H}_E$. For given $\mathbf{C}$, define the isometries $\tilde{\mathbf{V}}_\Phi$ and $\tilde{\mathbf{V}}_\Psi$ from $\mathbb{H}_A$ into $\mathbb{H}_B \otimes (\mathbb{H}_{E_1} \oplus \mathbb{H}_{E_2})$ by setting

$$\tilde{\mathbf{V}}_\Phi|\varphi\rangle_A = \mathbf{V}_\Phi|\varphi\rangle_A \oplus |0\rangle$$

and

$$\tilde{\mathbf{V}}_\Psi|\varphi\rangle_A = \big((\mathbf{I}_B \otimes \mathbf{C})\mathbf{V}_\Psi|\varphi\rangle_A\big) \oplus \big(\mathbf{I}_B \otimes \sqrt{\mathbf{I}_E - \mathbf{C}^*\mathbf{C}}\mathbf{V}_\Psi|\varphi\rangle_A\big)$$

for any $\varphi \in \mathbb{H}_A$, where we assume that the isometries $\mathbf{V}_\Phi$ and $\mathbf{V}_\Psi$ act from $\mathbb{H}_A$ to $\mathbb{H}_B \otimes \mathbb{H}_E^1$ and $\mathbb{H}_B \otimes \mathbb{H}_E^2$ correspondingly, while the contraction $\mathbf{C}$ acts from $\mathbb{H}_E^2$ to $\mathbb{H}_E^1$. It is easy to see that the isometries $\tilde{\mathbf{V}}_\Phi$ and $\tilde{\mathbf{V}}_\Psi$ form a common Stinespring representation for the channels Φ and Ψ with the required property. This proves that $\mathcal{M}(\Phi,\Psi) \subseteq \mathcal{N}(\Phi,\Psi)$.

To prove that $\mathcal{N}(\Phi,\Psi) \subseteq \mathcal{M}(\Phi,\Psi)$, take any common Stinespring representation for Φ and Ψ with the isometries $\tilde{\mathbf{V}}_\Phi$ and $\tilde{\mathbf{V}}_\Psi$ from $\mathbb{H}_A$ to $\mathbb{H}_B \otimes \mathbb{H}_{\tilde{E}}$. By Definition 5.2.2 and (5.4), there exist partial isometries $\mathbf{W}_\Phi$ and $\mathbf{W}_\Psi$ from $\mathbb{H}_E$ to $\mathbb{H}_{\tilde{E}}$ such that

$$\tilde{\mathbf{V}}_\Phi = (\mathbf{I}_B \otimes \mathbf{W}_\Phi)\mathbf{V}_\Phi \quad \text{and} \quad \tilde{\mathbf{V}}_\Psi = (\mathbf{I}_B \otimes \mathbf{W}_\Psi)\mathbf{V}_\Psi.$$

Therefore, $\tilde{\mathbf{V}}_\Phi^*\tilde{\mathbf{V}}_\Psi = \tilde{\mathbf{V}}_\Phi^*(\mathbf{I}_B \otimes \mathbf{W}_\Phi^*\mathbf{W}_\Psi)\mathbf{V}_\Psi \in \mathcal{M}(\Phi,\Psi)$, since $\|\mathbf{W}_\Phi^*\mathbf{W}_\Psi\|_\infty \leq 1$. This shows that $\mathcal{N}(\Phi,\Psi) \subseteq \mathcal{M}(\Phi,\Psi)$.

Since $\mathcal{N}(\Phi,\Psi) = \mathcal{M}(\Phi,\Psi)$, the infimum in (6.45) can be taken over the set $\mathcal{M}(\Phi,\Psi)$. This implies

$$\begin{aligned}\beta_E'(\Phi,\Psi) &= \inf_{\mathbf{C}\in\mathfrak{B}_1(\mathbb{H}_E)} \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} \sqrt{2 - 2\Re\big(\mathrm{tr}[\mathbf{V}_\Phi^*(\mathbf{I}_B \otimes \mathbf{C})\mathbf{V}_\Psi\rho]\big)} \\ &= \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} \inf_{\mathbf{C}\in\mathfrak{B}_1(\mathbb{H}_E)} \sqrt{2 - 2\Re\big(\mathrm{tr}[\mathbf{V}_\Phi^*(\mathbf{I}_B \otimes \mathbf{C})\mathbf{V}_\Psi\rho]\big)},\end{aligned} \tag{6.46}$$

where the possibility to change the order of the optimization follows from Ky Fan's well-known minimax theorem (see Simons [165]) and the σ-weak compactness of the closed unit ball $\mathfrak{B}_1(\mathbb{H}_E)$ of $\mathfrak{B}(\mathbb{H}_E)$ (by the Banach–Alaoglu Theorem 1.1.4). It is easy to see that

$$\begin{aligned}&\sup_{\mathbf{C}\in\mathfrak{B}_1(\mathbb{H}_E)} \Re\big(\mathrm{tr}[\mathbf{V}_\Phi^*(\mathbf{I}_B \otimes \mathbf{C})\mathbf{V}_\Psi\rho]\big) \\ &\quad = \sup_{\mathbf{C}\in\mathfrak{B}_1(\mathbb{H}_E)} \big|\langle(\mathbf{V}_\Phi \otimes \mathbf{I}_R)\varphi|\mathbf{I}_{BR} \otimes \mathbf{C}|\mathbf{V} \otimes \mathbf{I}_R\rangle_{AR}\big|,\end{aligned} \tag{6.47}$$

where R is any system and φ is a purification of ρ, i. e., a vector in $\mathbb{H}_{AR}$ such that $\mathrm{tr}_R[|\varphi\rangle_{AR}\langle\varphi|] = \rho$.

Since for any common Stinespring representation (6.44) and any system R, the vectors $(\mathbf{V}_\Phi \otimes \mathbf{I}_R)|\varphi\rangle_{AR}$ and $(\mathbf{V}_\Psi \otimes \mathbf{I}_R)|\varphi\rangle_{AR}$ in $\mathbb{H}_{BER}$ are purifications of the states

$(\Phi \otimes \mathbf{I}_R)(|\varphi\rangle_{AR}\langle\varphi|)$ and $(\mathbf{V}_\Psi \otimes \mathbf{I}_R)(|\varphi\rangle_{AR}\langle\varphi|)$ in $\mathcal{S}(\mathbb{H}_{BR})$, by using the relation $\mathcal{N}(\Phi,\Psi) = \mathcal{M}(\Phi,\Psi)$ proved earlier and Uhlmann's Theorem 6.1.5, it is easy to show that the quantity in the right-hand side, of (6.47) coincides with the fidelity of the states $(\Phi \otimes \mathbf{I}_R)(|\varphi\rangle_{AR}\langle\varphi|)$ and $(\Psi \otimes \mathbf{I}_R)(|\varphi\rangle_{AR}\langle\varphi|)$. This and (6.46) implies that $\beta'_E(\Phi,\Psi) = \beta_E(\Phi,\Psi)$. This proves the proposition. □

Proposition 6.4.6. *Let* $\mathbf{H}$ *be any* $\mathfrak{H}$*-operator defined on* $\mathbb{H}$*. For any given channels* Φ *and* Ψ*, the following properties hold:*
1. $\beta_E(\Phi,\Psi)$ *tends to* $\beta(\Phi,\Psi)$ *as* $E \to +\infty$*;*
2. *the function* $E \mapsto \beta_E(\Phi,\Psi)$ *is concave and nondecreasing on* $[E_0, +\infty[$.

If the operator $\mathbf{H}$ *satisfies condition* $\mathrm{tr}[e^{-\lambda\mathbf{H}}] < +\infty$ *for all* $\lambda > 0$*, then for any* $E > E_0$ *the metric* $\beta_E(\cdot,\cdot)$ *generates the strong (pointwise) convergence of quantum channels. That is,* $\lim_{n\to+\infty} \beta_E(\Phi_n,\Phi) = 0$ *if and only if* $\lim_{n\to+\infty} \|\Phi_n(\rho) - \Phi(\rho)\|_1 = 0$ *for all* $\rho \in \mathcal{S}(\mathbb{H}_A)$.

Proof. Property 1 and the monotonicity of β_E are obvious. To prove the concavity, take any $E_1, E_2 > E_0$. For arbitrary $\epsilon > 0$, there exist states ρ_1 and ρ_2 in $\mathcal{S}(\mathbb{H}_{AR})$ such that

$$\mathrm{tr}[\mathbf{H}\rho_i] \le E_i \quad \text{and} \quad \beta^2_{E_2}(\Phi,\Psi) \le \beta^2\big((\Phi \otimes \mathbf{I}_R)(\rho_i), (\Psi \otimes \mathbf{I}_R)(\rho_i)\big) + \epsilon, \quad i = 1, 2.$$

By invariance of the Bures distance (6.22) with respect to unitary transformation of both states, we may assume that $\mathrm{supp}(\mathrm{tr}_A[\rho_1]) \perp \mathrm{supp}(\mathrm{tr}_A[\rho_2])$, and hence, $\mathrm{supp}((\Upsilon_1 \otimes \mathbf{I}_R)(\rho_1)) \perp \mathrm{supp}((\Upsilon_2 \otimes \mathbf{I}_R)(\rho_2))$ for any channels Υ_1 and Υ_2. Thus,

$$\begin{aligned}
\beta^2_{\bar{E}}(\Phi,\Psi) &\ge \beta^2\big((\Phi \otimes \mathbf{I}_R)(\bar{\rho}_1), (\Psi \otimes \mathbf{I}_R)(\bar{\rho}_2)\big) \\
&= \frac{1}{2}\beta^2\big((\Phi \otimes \mathbf{I}_R)(\rho_1), (\Psi \otimes \mathbf{I}_R)(\rho_1)\big) + \frac{1}{2}\beta^2\big((\Phi \otimes \mathbf{I}_R)(\rho_2), (\Psi \otimes \mathbf{I}_R)(\rho_2)\big) \\
&\ge \frac{1}{2}\big(\beta^2_{E_1}(\Phi,\Psi) + \beta^2_{E_2}(\Phi,\Psi)\big) - \epsilon,
\end{aligned}$$

where $\bar{\rho} = \frac{1}{2}(\rho_1 + \rho_2)$ and $\bar{E} = \frac{1}{2}(E_1 + E_2)$. This implies concavity of the function $E \mapsto \beta^2_E(\Phi,\Psi)$. The concavity of the function $E \mapsto \beta_E(\Phi,\Psi)$ follows from the concavity and monotonicity of the function $\sqrt{x}$. Note that the operator $\mathbf{H}$ satisfies condition $\mathrm{tr}[e^{-\lambda\mathbf{H}}] < +\infty$ for all $\lambda > 0$ guarantees that the operator is unbounded and densely defined on $\mathbb{H}_A$ with a discrete spectrum of finite multiplicity. The last assertion of the proposition follows immediately. This proves the proposition. □

6.5 Approximations of quantum channels

Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be a sequence of (orthogonal) projections that converges strongly to the identity operator $\mathbf{I}_A$ on the Hilbert space $\mathbb{H}_A$.

In the following, we have an approximation result of the infinite-dimensional quantum channel via finite-dimensional projections. The result is due originally to Shirokov and Holevo [158].

Proposition 6.5.1. *Let* $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ *be an extended quantum channel from input system A to output system B, and let* $(\mathbf{P}_n)_{n=1}^{+\infty}$ *be an increasing sequence of finite-dimensional projections on* $\mathbb{H}_A$ *that converges strongly to the identity operator* $\mathbf{I}_A$. *Then there is a family* $(\Phi_n)_{n=1}^{+\infty}$ *of completely positive maps such that* Φ_n *is trace preserving on* $\mathbf{P}_n(\mathbb{H}_A)$ *and* $\Phi_n(\mathbf{A}) \to \Phi(\mathbf{A})$ *uniformly for all* $\mathbf{A} \in \mathfrak{T}_+(\mathbb{H}_A)$.

Proof. First note that, for each $\mathbf{A} \in \mathfrak{B}(\mathbb{H}_A)$, $\mathbf{A}_n \equiv \mathbf{P}_n\mathbf{A}\mathbf{P}_n^* \to \mathbf{A}$ in operator norm $\|\cdot\|_\infty$, when $n \to +\infty$. We can write $\mathbf{A} - \mathbf{A}_n = (\mathbf{A} - \mathbf{A}_n)_+ - (\mathbf{A} - \mathbf{A}_n)_-$, where $(\mathbf{A} - \mathbf{A}_n)_+$ and $(\mathbf{A} - \mathbf{A}_n)_-$ are in $\mathfrak{T}_+(\mathbb{H}_A)$ and $(\mathbf{A} - \mathbf{A}_n)_+(\mathbf{A} - \mathbf{A}_n)_- = (\mathbf{A} - \mathbf{A}_n)_-(\mathbf{A} - \mathbf{A}_n)_+ = \mathbf{0}$. Note that $(\mathbf{A} - \mathbf{A}_n)_+$ and $(\mathbf{A} - \mathbf{A}_n)_-$ are, respectively, the positive and negative part of $\mathbf{A} - \mathbf{A}_n$. Obviously,

$$\Phi\big((\mathbf{A} - \mathbf{A}_n)_+\big), \Phi\big((\mathbf{A} - \mathbf{A}_n)_-\big) \in \mathfrak{T}_+(\mathbb{H}_B)$$

and

$$\lim_{n\to+\infty} \operatorname{tr}\big[\Phi\big((\mathbf{A} - \mathbf{A}_n)_+\big)\big] = \lim_{n\to+\infty} \operatorname{tr}\big[\Phi\big((\mathbf{A} - \mathbf{A}_n)_-\big)\big] = 0.$$

Consequently,

$$\lim_{n\to+\infty} \big\|\Phi\big((\mathbf{A} - \mathbf{A}_n)_+\big)\big\|_1 = \lim_{n\to+\infty} \big\|\Phi\big((\mathbf{A} - \mathbf{A}_n)_-\big)\big\|_1 = 0$$

and

$$\lim_{n\to+\infty} \big\|\Phi(\mathbf{A} - \mathbf{A}_n)\big\|_1 = 0.$$

For each $n \in \mathbb{N}$, define $\Psi_n : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ through $\Psi_n(\mathbf{A}) = \mathbf{P}_n\Phi(\mathbf{A}_n)\mathbf{P}_n^*$. Then

$$\begin{aligned}
&\big\|\Phi(\mathbf{A}) - \Psi_n(\mathbf{A})\big\|_1 \\
&\quad\le \big\|\Phi(\mathbf{A}) - \mathbf{P}_n\Phi(\mathbf{A})\mathbf{P}_n^*\big\|_1 + \big\|\mathbf{P}_n\Phi(\mathbf{A})\mathbf{P}_n^* - \mathbf{P}_n\Phi(\mathbf{A}_n)\mathbf{P}_n^*\big\|_1 \\
&\quad= \big\|\Phi(\mathbf{A}) - \mathbf{P}_n\Phi(\mathbf{A})\mathbf{P}_n^*\big\|_1 + \big\|\mathbf{P}_n\big(\Phi(\mathbf{A}) - \Phi(\mathbf{A}_n)\big)\mathbf{P}_n^*\big\|_1 \\
&\quad\le \big\|\Phi(\mathbf{A}) - \Phi(\mathbf{A}_n)\big\|_1 + \big\|\Phi(\mathbf{A}) - \Phi(\mathbf{A}_n)\big\|_1 \to 0 \quad \text{as } n \to +\infty.
\end{aligned}$$

Define Φ_n by $\Phi_n(\mathbf{A}) = \Psi_n(\mathbf{A}) + \mathbf{V}_n\mathbf{A}\mathbf{V}_n^*$, where $\mathbf{V}_n = \sqrt{\mathbf{P}_n - \Psi_n^+(\mathbf{I})}$. Then $\|\Phi(\mathbf{A}) - \Phi(\mathbf{A}_n)\|_1 \to 0$ and $\operatorname{tr}[\Phi_n(\mathbf{A})] = \operatorname{tr}[\mathbf{A}]$ for all $\mathbf{A}$ such that $\mathbf{A} = \mathbf{A}_n$. This proves the proposition. ☐

6.6 Convergences of channels

In the following, let $(\Phi_n)_{n=0}^{+\infty}$ be a sequence of quantum operations or quantum channels from system A to system B. This section explores three different types of convergences (namely, uniform convergence, strong* convergence, and strong convergence) of the sequence $(\Phi_n)_{n=1}^{+\infty}$ to Φ_0.

6.6.1 Strong and uniform convergence

Definition 6.6.1 (Strong convergence of quantum operations). The sequence of quantum operations $(\Phi_n)_{n=1}^{+\infty}$ in $\mathfrak{QO}(A,B)$ is said to converge strongly to operation $\Phi_0 \in \mathfrak{QO}(A,B)$ if

$$\lim_{n\to\infty} \|\Phi_n(\rho) - \Phi_0(\rho)\|_1 = 0, \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A). \tag{6.48}$$

In this case, we write $(s)\lim_{n\to+\infty} \Phi_n = \Phi_0$.

The strong convergence of quantum operations is generated by the family of seminorms $\Phi \mapsto \|\Phi\|_1$.

Definition 6.6.2 (Uniform convergence of quantum operations). The sequence of quantum operations $(\Phi_n)_{n=1}^{+\infty}$ in $\mathfrak{QO}(A,B)$ is said to converge to $\Phi_0 \in \mathfrak{QO}(A,B)$ uniformly if

$$\lim_{n\to+\infty} \sup_{\rho\in\mathcal{S}(\mathbb{H}_A)} \|\Phi_n(\rho) - \Phi_0(\rho)\|_1 = 0. \tag{6.49}$$

In this case, we write $(u)\lim_{n\to+\infty} \Phi_n = \Phi_0$.

It is clear that if the sequence $(\Phi_n)_{n=1}^{+\infty}$ quantum operations converges uniformly to Φ_0 then it converges strongly to Φ_0. However, we have the following result due originally to Shirokov and Holevo [158].

Lemma 6.6.3. *The strong convergence topology on $\mathfrak{QO}(A,B)$ coincides with the topology of uniform convergence on compact subsets of $\mathcal{S}(\mathbb{H}_A)$.*

Proof. Let $\Phi \in \mathfrak{QO}(A,B)$. Then

$$\|\Phi\|_{1,\infty} := \sup_{\rho\in\mathcal{S}(\mathbb{H}_A)} \frac{\|\Phi(\rho)\|_1}{\|\rho\|_1} \le \sup_{\rho\in\mathcal{S}(\mathbb{H}_A)} \frac{\|\rho\|_1}{\|\rho\|_1} = 1,$$

where $\|\rho\|_1$ denotes the trace norm of ρ in the Hilbert space $\mathbb{H}_A$ and $\|\Phi(\rho)\|_1$ denotes the trace norm of channel output $\Phi(\rho)$ in the Hilbert space $\mathbb{H}_B$. It is easily seen that the topology of strong convergence on the space $\mathfrak{QO}(A,B)$ coincides with the topology of uniform convergence on $\mathfrak{QO}(A,B)$ that are restricted to compact subsets of $\mathcal{S}(\mathbb{H}_A)$.

This is because the supremum in (6.49) is attained in the compact subset of $\mathcal{S}(\mathbb{H}_A)$. This proves the lemma. □

Suppose that the sequence $(\Phi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A,B)$ converges strongly to $\Phi_0 \in \mathfrak{QC}(A,B)$. Then its corresponding sequence of dual channels $(\Phi_n^*)_{n=1}^{+\infty} \subset \mathfrak{QC}^*(B,A)$ has the following convergence relationship to $\Phi_0^* \in \mathfrak{QC}^*(B,A)$:

$$(w)\lim_{n\to+\infty} \Phi_n^*(\mathbf{B}) = \Phi_0^*(\mathbf{B}), \quad \forall \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B).$$

That is, $(\Phi_n^*(\mathbf{B}))_{n=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H}_A)$ converges to $\Phi_0^*(\mathbf{B}) \in \mathfrak{B}(\mathbb{H}_A)$ in weak operator topology (see Definition 2.1.2 for the definition of weak operator topology). In this case,

$$\lim_{n\to+\infty} \langle \phi, \Phi_n^*(\mathbf{B})\psi\rangle_{\mathbb{H}_A} = \langle \phi, \Phi_0^*(\mathbf{B})\psi\rangle_{\mathbb{H}_A}, \quad \forall \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B) \text{ and } \forall \phi,\psi \in \mathbb{H}_A.$$

By noting that the set $\mathcal{S}(\mathbb{H}_A)$ involved in (6.48) can be replaced by its subset consisting of pure states, it is easy to show that the strong convergence of a sequence $(\Phi_n)_{n=1}^{+\infty}$ of quantum channels to a channel Φ_0 means, in the Heisenberg picture, that

$$(w)\lim_{n\to+\infty} \Phi_n^*(\mathbf{B}) = \Phi_0^*(\mathbf{B}), \quad \forall \mathbf{B} \in \mathfrak{B}(\mathbb{H}_B), \tag{6.50}$$

where $(w)\lim_{n\to+\infty}(\cdot)$ denotes the limit in the weak operator topology in $\mathfrak{B}(\mathbb{H}_A)$.

Lemma 6.6.4. *The strong convergence of a sequence $(\Phi_n)_{n=1}^{+\infty}$ in $\mathfrak{QC}(A,B)$ to a channel Φ_0 implies strong convergence of the sequence $(\Phi_n \otimes \mathfrak{I}_R)_{n=1}^{+\infty}$ to the channel $\Phi_0 \otimes \mathfrak{I}_R$, where R is any system.*

Proof. Assume that the sequence quantum channels $(\Phi_n)_{n=1}^{+\infty}$ converges to Φ_0 strongly. For any $\rho_{AR} \in \mathcal{S}(\mathbb{H}_{AR})$, we have

$$\begin{aligned}
&\lim_{n\to+\infty} \big\|(\Phi_n \otimes \mathfrak{I}_R)(\rho_{AR}) - (\Phi_0 \otimes \mathfrak{I}_R)(\rho_{AR})\big\|_1 \\
&\quad = \lim_{n\to+\infty} \operatorname{tr}\big[\big|(\Phi_n \otimes \mathfrak{I}_R)(\rho_{AR}) - (\Phi_0 \otimes \mathfrak{I}_R)(\rho_{AR})\big|\big] \\
&\quad = \lim_{n\to+\infty} \operatorname{tr}\big[\operatorname{tr}_R\big[|(\Phi_n \otimes \mathfrak{I}_R)(\rho_{AR}) - (\Phi_0 \otimes \mathfrak{I}_R)(\rho_{AR})|\big]\big] \\
&\quad = \lim_{n\to+\infty} \operatorname{tr}\big[|\Phi_n(\rho_A) - \Phi_0(\rho_A)|\big] \\
&\quad = \lim_{n\to+\infty} \|\Phi_n(\rho_A) - \Phi_0(\rho_A)\|_1 = 0,
\end{aligned}$$

where $\rho_A = \operatorname{tr}_R[\rho_{AR}]$. In the above, we have used the fact that $\operatorname{tr}[\operatorname{tr}_R[(\Phi_n \otimes \mathfrak{I}_R)(\rho_{AR})]] = \operatorname{tr}[\Phi_n(\rho_A)]$ by (2.11). This proves the lemma. □

The diamond-norm distance between two quantum channels is widely used as a measure of distinguishability between these channels. But the topology (convergence) generated by the diamond-norm distance on the set of infinite-dimensional quantum channels is too strong for analysis of real variations of such channels. We offer the following example to accentuate this comment.

Example 6.1. Let $(\mathbf{U}_n)_{n=1}^{+\infty}$ be any sequence of unitary operators strongly converging to the unit operator $\mathbf{I}_{\mathbb{H}}$ but not converging to $\mathbf{I}_{\mathbb{H}}$ in the operator norm. That is,

$$\lim_{n\to+\infty} \|\mathbf{U}_n\psi - \psi\|_{\mathbb{H}} = 0, \quad \forall\psi \in \mathbb{H} \quad \text{but} \lim_{n\to+\infty} \|\mathbf{U}_n - \mathbf{I}_{\mathbb{H}}\|_\infty \neq 0.$$

Consider the sequence of channels $(\Phi_n)_{n=1}^{+\infty}$ from $\mathcal{S}(\mathbb{H})$ into itself defined by $\Phi_n(\rho) = \mathbf{U}_n\rho\mathbf{U}_n^*$ for $n = 1, 2, \ldots$. Then this sequence of channels does not converge to the identity channel $\mathfrak{I}_{\mathbb{H}}$ with respect to the diamond-norm distance. This is because

$$\begin{aligned}
&\lim_{n\to+\infty} \|\Phi_n - \mathfrak{I}_{\mathbb{H}}\|_\diamond \\
&\quad = \lim_{n\to+\infty}\Big(\sup_{\rho_{AR}\in\mathcal{S}(\mathbb{H}_{AR})} \big\|(\Phi_n \otimes \mathfrak{I}_R - \mathfrak{I}_{\mathbb{H}} \otimes \mathfrak{I}_R)(\rho_{AR})\big\|_1\Big) \\
&\quad = \lim_{n\to+\infty}\Big(\sup_{\rho_A\in\mathcal{S}(\mathbb{H}_A)} \big\|(\Phi_n - \mathfrak{I}_{\mathbb{H}})(\rho_A)\big\|_1\Big) \quad (\text{where } \rho_A = \mathrm{tr}_R[\rho_{AR}]) \\
&= \lim_{n\to+\infty}\Big(\sup_{\rho_A\in\mathcal{S}(\mathbb{H}_A)} \big\|\mathbf{U}_n\rho_A\mathbf{U}_n^* - \rho_A\big\|_1\Big) \neq 0.
\end{aligned}$$

In general, the closeness of two quantum channels in the diamond-norm distance means, by Theorem 6.3.2, the operator-norm closeness of the corresponding Stinespring isometries. So, if we use the diamond-norm distance then we take into account only such perturbations of a channel that corresponds to uniform deformations of the Stinespring isometry (i. e., deformations with small operator norm). As a result, there exist but not shown here quantum channels with close physical parameters (quantum limited attenuators) having the diamond-norm distance equal to 2.

The following equivalency of energy-constrained diamond norm and strong convergence of channels is due originally to Shirokov [151].

Proposition 6.6.5. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator defined on* $\mathbb{H}_A$ *and let* $(\Phi_n)_{n=0}^{+\infty} \subset \mathfrak{QC}(A, B)$. *Then, for any* $E \geq E_0$,

$$\lim_{n\to+\infty} \|\Phi_n - \Phi_0\|_\diamond^E = 0 \Leftrightarrow \lim_{n\to+\infty} \Phi_n(\rho) = \Phi_0(\rho), \quad \forall\rho \in \mathcal{S}(\mathbb{H}_A). \tag{6.51}$$

Proof. $(\Rightarrow)$ Recall from Theorem 3.2.5 that $\mathcal{K}_{\mathbf{H}}(E)$ is a compact subset of $\mathcal{S}(\mathbb{H}_A)$ and that $\overline{\bigcup_{E\geq E_0} \mathcal{K}_{\mathbf{H}}(E)}^{\|\cdot\|_1} = \mathcal{S}((\mathbb{H}_A)$. Hence, since the operator norm of $\Phi_n - \Phi_0$ for all n is bounded, it suffices to show that

$$\lim_{n\to+\infty} \Phi_n(\rho) = \Phi_0(\rho), \quad \forall\rho \in \bigcup_{E\geq E_0} \mathcal{K}_{\mathbf{H}}(E),$$

provided that $\lim_{n\to+\infty} \|\Phi_n - \Phi_0\|_\diamond^E = 0$. Let ρ be a state in $\bigcup_{E\geq E_0} \mathcal{K}_{\mathbf{H}}(E)$ and $\sigma \in \mathcal{K}_{\mathbf{H}}(E)$. Then, for sufficiently small $p > 0$ the energy of the state $\rho_p = (1 - p)\sigma + p\rho$ does not exceed E. Hence, $\lim_{n\to+\infty} \|\Phi_n - \Phi_0\|_\diamond^E = 0$ implies that

$$\lim_{n\to+\infty}\left\|\Phi_n(\rho_p)-\Phi_0(\rho_p)\right\|_1=0 \quad \text{and} \quad \lim_{n\to+\infty}\left\|\Phi_n(\sigma)-\Phi_0(\sigma)\right\|_1=0.$$

It follows that $\lim_{n\to+\infty}\|\Phi_n(\rho)-\Phi_0(\rho)\|_1=0$ for all $\rho\in\mathcal{S}(\mathbb{H}_A)$.

($\Leftarrow$) Since $\mathbf{H}$ is an $\mathfrak{H}$-operator, we can write $\mathbf{H}$ as $\mathbf{H}=\sum_{k=0}^{+\infty}E_k|\tau_k\rangle_A\langle\tau_k|$, where $\{|\tau_k\rangle_A\}:=\{|\tau_k\rangle_{\mathbb{H}_A}\}$ is the orthonormal basis of eigenvectors of $\mathbf{H}$ corresponding to the nondecreasing sequence of eigenvalues $(E_k)_{k=0}^{+\infty}$ that tend to $+\infty$. Let $\mathbf{P}_n=\sum_{k=0}^{n-1}|\tau_k\rangle_A\langle\tau_k|$ be the projector on the subspace $\mathbb{H}_n$ spanned by the set of vectors $\{|\tau_0\rangle_A,\dots,|\tau_{n-1}\rangle_A\}$. Consider the family of seminorms

$$q_n(\Phi):=\sup_{\rho\in\mathcal{S}(\mathbb{H}_n\otimes\mathbb{H}_R)}\left\|(\Phi\otimes\mathbf{I}_R)(\rho)\right\|_1,\quad \forall n\in\mathbb{N} \tag{6.52}$$

on the set of all linear completely bounded maps from $\mathfrak{T}(\mathbb{H}_A)$ to $\mathfrak{T}(\mathbb{H}_B)$. Note that the system R in (6.52) may be n-dimensional. Indeed, by convexity of the trace norm the supremum in (6.52) can be taken over pure states ρ in $\mathcal{S}(\mathbb{H}_n\otimes\mathbb{H}_R)$. Since the marginal state ρ_R of any such pure state ρ has rank $\le n$, by applying local unitary transformation of the system R we can put all these states into the set $\mathcal{S}(\mathbb{H}_n\otimes\mathbb{H}_{R,n})$, where $\mathbb{H}_{R,n}$ is any n-dimensional subspace of $\mathbb{H}_R$. Let $(\Phi_k)_{k=1}^{+\infty}$ be a sequence of channels strongly converging to a channel Φ_0. By Definition 6.6.2 and Lemma 6.6.4, this implies that $\sup_{\rho\in\mathcal{K}}\|(\Phi_k-\Phi_0)\otimes\mathbf{I}_R)(\rho)\|_1$ tends to zero for any compact subset $\mathcal{K}$ of $\mathcal{S}(\mathbb{H}_{AR})$. Since $\dim(\mathbb{H}_R)=n$, the set $\mathcal{S}(\mathbb{H}_n\otimes\mathbb{H}_R)$ is compact and we conclude that

$$\lim_{k\to+\infty}q_n(\Phi_k-\Phi_0)=0,\quad \forall n\in\mathbb{N}. \tag{6.53}$$

Let $E\ge E_0$ and ρ be a state in $\mathcal{S}(\mathbb{H}_{AR})$ such that $\mathrm{tr}[\mathbf{H}\rho_A]\le E$, where $\rho_A=\mathrm{tr}_R[\rho]$. Then the state $\rho_n=(1-r_n)^{-1}(\mathbf{P}_n\otimes\mathbf{I}_R)\rho(\mathbf{P}_n\otimes\mathbf{I}_R)$ (where $r_n=\mathrm{tr}[(\mathbf{I}_A-\mathbf{P}_n)\rho_A]$) belongs to the set $\mathcal{S}(\mathbb{H}_n\otimes\mathbb{H}_R)$. By using the inequality,

$$\left\|((\mathbf{I}_A-\mathbf{P}_n)\otimes\mathbf{I}_R)\rho(\mathbf{P}_n-\mathbf{I}_R)\right\|_1\le\mathrm{tr}[((\mathbf{I}_A-\mathbf{P}_n)\otimes\mathbf{I}_R)\rho]=\sqrt{r_n}$$

(which can be easily proved via the operator version of Cauchy–Schwarz inequality) and by noting that $\mathrm{tr}[\mathbf{H}\rho_A]\le E$ implies $r_n\le E/E_n$, we obtain

$$\|\rho-\rho_n\|_1\le 2\left\|((\mathbf{I}_A-\mathbf{P}_n)\otimes\mathbf{I}_R)\rho(\mathbf{P}_n\otimes\mathbf{I}_R)\right\|_1+2r_n\le 4\sqrt{r_n}\le 4\sqrt{E/E_n}.$$

It follows that

$$\|\Phi_k-\Phi_0\|_\diamond^E\le q_n(\Phi_k-\Phi_0)+8\sqrt{E/E_n}.$$

Since $E_n\to+\infty$ as $n\to+\infty$, this inequality and (6.53) show that $\|\Phi_n-\Phi_0\|_\diamond^E\to 0$ as $n\to+\infty$. This proves the proposition. □

6.6.2 Strong* convergence of channels

Definition 6.6.6. A sequence of quantum channels $(\Phi_n)_{n=1}^{+\infty}$ in $\mathfrak{QC}(A, B)$ is said to be strongly* converges to Φ_0 in $\mathfrak{QC}(A, B)$ if

$$\lim_{n\to+\infty}\|\Phi_n(\rho) - \Phi_0(\rho)\|_1 = 0, \quad \forall\rho \in \mathcal{S}(\mathbb{H}_A) \tag{6.54}$$

and

$$\lim_{n\to+\infty}\|\Phi_n^*(\mathbf{B}) - \Phi_0^*(\mathbf{B})\|_\infty = 0, \quad \forall\mathbf{B} \in \mathfrak{B}(\mathbb{H}_B). \tag{6.55}$$

In this case, we write $(s^*)\lim_{n\to+\infty}\Phi_n = \Phi_0$.

By using (6.55) as the simplest definition of the strong* convergence, it is easy to show (see the following proposition due originally to Shirokov [153]) that this convergence is preserved under basic manipulations with quantum channels.

Proposition 6.6.7. *We have the following:*

1. *Let* $(\Phi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A, B)$ *and* $(\Psi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(C, D)$ *be sequences of quantum channels that* strongly* *converge to quantum channels* $\Phi_0 \in \mathfrak{QC}(A, B)$ *and* $\Psi_0 \in \mathfrak{QC}(C, D)$, *respectively. Then (i) the sequence* $(\Phi_n \otimes \Psi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(AC, BD)$ *of channels from AC to BD* strongly* *converges to the channel* $\Phi_0 \otimes \Psi_0 \in \mathfrak{QC}(AC, BD)$; *(ii) If* $B = C$, *then the sequence* $(\Psi_n \circ \Phi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A, D)$ *of channels from A to D* strongly* *converges to the channel* $\Psi_0 \circ \Phi_0 \in \mathfrak{QC}(A, D)$; *and (iii) if* $(\Phi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A, B)$ strongly* *converges to* $\Phi_0 \in \mathfrak{QC}(A, B)$ *then its corresponding sequence of complementary channels* $(\hat{\Phi}_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A, E)$ strongly* *converges to the complementary channel* $\hat{\Phi}_0 \in \mathfrak{QC}(A, E)$ *of* $\Phi_0 \in (A, B)$.
2. *If a sequence of channels* $(\Phi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A, B)$ *from A to B* strongly* *converging to a channel* $\Phi_0 \in \mathfrak{QC}(A, B)$ *then there exists a sequence* $(\Psi_n)_{n=1}^{+\infty}$ *of quantum channels from A to some system E that* strongly* *converges to a channel* $\Psi_0 \in \mathfrak{QC}(A, E)$ *such that* $\Psi_n = \hat{\Phi}_n$ *for all* $n \geq 0$, *where* $\hat{\Phi}_n$ *is the complementary channel of* Φ_n *as defined in Definition* 5.7.1.

Proof. The proof of the proposition can easily be achieved by using the definition of strong* convergence (see equation (6.55)) on the manipulation of tensor product and composition of two sequences of strong* convergent quantum channels. The details of the proof is omitted here. □

Strong* convergence of quantum channels is stronger than the strong convergence. The difference between these types of convergence is best seen on the Heisenberg picture: as mentioned earlier, the strong convergence of a sequence $(\Phi_n)_{n=1}^{+\infty}$ to a channel Φ_0 can be defined as

$$(w)\lim_{n\to+\infty}\Phi_n^*(\mathbf{B}) = \Phi_0^*(\mathbf{B}), \quad \forall\mathbf{B} \in \mathfrak{B}(\mathbb{H}_B)$$

(the limit in the weak operator topology) while the strong* convergence of this sequence means that

$$(s)\lim_{n\to+\infty}\Phi_n^*(\mathbf{B})=\Phi_0^*(\mathbf{B}),\quad \forall \mathbf{B}\in\mathfrak{B}(\mathbb{H}_B).$$

While strong* convergence of quantum channels is stronger than strong convergence, the following example shows, however, that the strong* convergence is substantially weaker than the uniform (diamond-norm) convergence.

Example 6.2. Let Φ be an arbitrary quantum channel from A to B, and let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be a sequence of finite rank projectors in $\mathfrak{B}(\mathbb{H}_B)$ strongly converging to the identity operator $\mathbf{I}_B := \mathbf{I}_{\mathbb{H}_B}$. Let

$$\Phi_n(\rho)=\mathbf{P}_n\Phi(\rho)\mathbf{P}_n+\big(\mathrm{tr}[(\mathbf{I}_B-\mathbf{P}_n)\Phi(\rho)]\big)\sigma,\quad \forall n,$$

where σ is a given state in $\mathcal{S}(\mathbb{H}_B)$. Sequences of this type are used in Holevo and Shirokov [158] for approximation of infinite-dimensional quantum channels by channels with the finite-dimensional output system (see Proposition 6.5.1). We claim that

(A) the sequence $(\Phi_n)_{n=1}^{+\infty}$ strongly* converges to the channel Φ; and

(B) the sequence $(\Phi_n)_{n=1}^{+\infty}$ does not converge uniformly to Φ in general.

We first prove (A) by using the duality relation

$$\mathrm{tr}[\Phi_n(\rho)\mathbf{B}]=\mathrm{tr}[\rho\Phi_n^*(\mathbf{B})],\quad \forall\rho\in\mathcal{S}(\mathbb{H}_A)\text{ and }\forall\mathbf{B}\in\mathfrak{B}(\mathbb{H}_B).$$

For all $\mathbf{B}\in\mathfrak{B}(\mathbb{H}_B)$ and for all $\rho\in\mathcal{S}(\mathbb{H}_A)$, we have

$$\begin{aligned}\mathrm{tr}[\rho\Phi_n^*(\mathbf{B})]&=\mathrm{tr}[\Phi_n(\rho)\mathbf{B}]\\&=\mathrm{tr}[\mathbf{P}_n\Phi(\rho)\mathbf{P}_n\mathbf{B}]+\mathrm{tr}[(\mathrm{tr}_B[(\mathbf{I}_B-\mathbf{P}_n)\Phi(\rho)])\sigma\mathbf{B}]\\&=\mathbf{P}_n\,\mathrm{tr}[\Phi(\rho)\mathbf{B}]\mathbf{P}_n+\mathrm{tr}[(\mathbf{I}_B-\mathbf{P}_n)\Phi(\rho)]\,\mathrm{tr}_B[\mathbf{B}\sigma]\\&=\mathbf{P}_n\,\mathrm{tr}[\rho\Phi^*(\mathbf{B})]\mathbf{P}_n+\mathrm{tr}[\rho\Phi^*((\mathbf{I}_B-\mathbf{P}_n)\,\mathrm{tr}_B[\mathbf{B}\sigma])]\\&=\mathrm{tr}[\rho\Phi^*(\mathbf{P}_n\mathbf{B}\mathbf{P}_n)]+\mathrm{tr}[\rho\Phi^*((\mathbf{I}_B-\mathbf{P}_n)\,\mathrm{tr}_B[\mathbf{B}\sigma])]\\&=\mathrm{tr}[\rho\Phi^*(\mathbf{P}_n\mathbf{B}\mathbf{P}_n+(\mathbf{I}_B-\mathbf{P}_n)\,\mathrm{tr}[\mathbf{B}\sigma])].\end{aligned}$$

This shows that $\Phi_n^*(\mathbf{B})=\Phi^*(\mathbf{P}_n\mathbf{B}\mathbf{P}_n+(\mathbf{I}_B-\mathbf{P}_n)\,\mathrm{tr}[\mathbf{B}\sigma])$. Since the map $\mathbf{B}\mapsto\Phi^*(\mathbf{B})$ is continuous with respect to the strong operator topology, we have for all $\mathbf{B}\in\mathfrak{B}(\mathbb{H}_B)$

$$\begin{aligned}(s)\lim_{n\to+\infty}\Phi_n^*(\mathbf{B})&=(s)\lim_{n\to+\infty}\Phi^*\big(\mathbf{P}_n\mathbf{B}\mathbf{P}_n+(\mathbf{I}_B-\mathbf{P}_n)\,\mathrm{tr}[\mathbf{B}\sigma]\big)\\&=(s)\lim_{n\to+\infty}\Phi^*(\mathbf{P}_n\mathbf{B}\mathbf{P}_n)+(s)\lim_{n\to\infty}\Phi^*\big((\mathbf{I_B}-\mathbf{P}_n)\,\mathrm{tr}[\mathbf{B}\sigma]\big)=\Phi^*(\mathbf{B}),\end{aligned}$$

due to the fact that

$$(s)\lim_{n\to+\infty}\Phi^*(\mathbf{P}_n\mathbf{B}\mathbf{P}_n) = \Phi^*\Big((s)\lim_{n\to\infty}(\mathbf{P}_n\mathbf{B}\mathbf{P}_n)\Big) = \Phi^*(\mathbf{B})$$

and

$$\begin{aligned}(s)&\lim_{n\to+\infty}\Phi^*((\mathbf{I}_B-\mathbf{P}_n)\operatorname{tr}[\mathbf{B}\sigma])\\ &=\Phi^*\Big((s)\lim_{n\to+\infty}(\mathbf{I}_\mathbf{B}-\mathbf{P}_n)\operatorname{tr}[\mathbf{B}\sigma]\Big)=\Phi^*(\mathbf{0})=\mathbf{0}.\end{aligned}$$

Therefore, the sequence $(\Phi_n)_{n=1}^{+\infty}$ strongly* converges to the channel Φ.

(B). The sequence $(\Phi_n)_{n=1}^{+\infty}$, however, does not converge uniformly to Φ in general. This is because one can consider the case where $A = B$ and $\Phi = \mathfrak{I}_A := \mathfrak{I}_{\mathbb{H}_A}$ (the identity channel on $\mathbb{H}_A = \mathbb{H}_B = \mathbb{H}$). In this case, $\Phi_n(\rho) = \mathbf{P}_n\rho\mathbf{P}_n + \operatorname{tr}[(\mathbf{I}_\mathbf{B} - \mathbf{P}_n)\rho]\sigma$ and it does not converge uniformly to $\Phi = \mathfrak{I}_A$, because

$$\begin{aligned}&\lim_{n\to+\infty}\sup_{\rho\in\mathcal{S}(\mathbb{H})}\|\Phi_n(\rho)-\Phi(\rho)\|_1\\ &=\lim_{n\to+\infty}\sup_{\rho\in\mathcal{S}(\mathbb{H})}\operatorname{tr}\big[\mathbf{P}_n\rho\mathbf{P}_n+\operatorname{tr}\big[(\mathbf{I}-\mathbf{P}_n)\rho\big]\sigma-\rho|\big]\neq 0\quad(\text{by choosing }\sigma=\rho)\end{aligned}$$

Therefore, the sequence $(\Phi_n)_{n=1}^{+\infty}$ does not converge uniformly to $\Phi = \mathfrak{I}$.

7 von Neumann entropy

This chapter deals with the relevant concepts and properties of von Neumann entropy of input quantum states before they are transmitted through the quantum channels from the sender to the receiver. In particular, continuity and/or discontinuity of von Neumann entropy on the whole space of $\mathcal{S}(\mathbb{H})$ or on certain closed convex subsets of $\mathcal{S}(\mathbb{H})$ will be explored. These properties play an important role in developing and defining channel capacities in later chapters. The topics represented here are sufficient enough for the study of infinite-dimensional quantum information with memory. For other topics of quantum entropies not touched in this book, the readers are referred to the paper by Wehrl [175] and the book by Ohya and Petz [121] for further reading.

We first study the continuity/discontinuity properties of von Neumann and relative entropies on the entire $\mathcal{S}(\mathbb{H})$. While these entropies are only lower semicontinuous on $\mathcal{S}(\mathbb{H})$, its restriction to some closed subsets or $\mathcal{S}(\mathbb{H})$ reveals many important and interesting behaviors.

7.1 von Neumann entropy on $\mathcal{S}(\mathbb{H})$

As a motivation for extension of classical entropies to quantum entropies and for comparison between the classical and quantum information theory, we first review the concept of classical Shannon entropy below.

In classical communication scenario, Alice, the sender, intends to transmit a message to Bob, the receiver, via a noisy classical channel. To transmit the message that consists of several letters, Alice first chooses a letter x from an alphabet Λ_X at random according to a random variable X with distribution $p_X(x)$ and encodes the selected letter into a codeword before it is transmitted through the channel. Due to noisiness of the channel, the receiver instead of receiving exactly what the sender had sent, he receives an alphabet y from the sample space of alphabets Λ_Y generated by the random variable Y. One of the goals in communication via a noisy classical channel is to determine $p_X(x)$, the probability mass function of the random variable X, via $p_{X|Y}(x|y)$, the conditional probability mass function of X given observation Y made by the receiver. The quantity that measures the receiver's uncertainty about X before learning it is the Shannon entropy denoted by $S(X)$ and defined by

$$S(X) = -\sum_{x\in\Lambda_X} p_X(x)\log(p_X(x)), \tag{7.1}$$

where log denotes the logarithmic function based 2 due to the fact that the messages are digitized in classical binary bit. Therefore, the entropy $S(X)$ can be viewed as the receiver's expected information gain, in the unit of bits, upon learning the outcome

https://doi.org/10.1515/9783110788105-007

of X. Shannon's noiseless coding theorem (see Shannon [140]) makes this interpretation precise by proving that the sender needs to send bits at a rate $S(X)$ in order for receiver to be able to decode a compressed message.

For communication through quantum channels, Shannon's entropy is extended to van Neumann entropy $H(\rho)$ for $\rho \in \mathcal{S}(\mathbb{H})$. For comparison purposes, the von Neumann entropy is defined in Definition 7.1.1 for finite-dimensional Hilbert space $\mathbb{H}$ and in Definition 7.1.2 for infinite-dimensional $\mathbb{H}$, which reduces to Definition 7.1.1 in finite dimensions. The definition of relative entropy is given in Definition 8.1.3.

In the following, we let the underlined quantum system be represented by a separable complex Hilbert space $\mathbb{H}$ in finite or infinite dimensions. Recall that the operator $\mathbf{A}$ on Hilbert space $\mathbb{H}$ is said to be a finite rank operator if the dimension of its range is a finite-dimensional Hilbert space. Properties of finite rank map/operator can be found in Section 1.7.

To generalize the concept of classical Shannon entropy directly to the content of quantum communication, we first define the von Neumann entropy $H(\mathbf{A})$ for positive finite rank operator $\mathbf{A}$ on the Hilbert space $\mathbb{H}$ below.

Definition 7.1.1. Let $\mathbf{A}$ be a positive finite rank operator on the separable complex Hilbert space $\mathbb{H}$. The von Neumann entropy $H(\mathbf{A})$ of $\mathbf{A}$ is defined by

$$H(\mathbf{A}) = \mathrm{tr}[\eta(\mathbf{A})], \tag{7.2}$$

where

$$\eta(x) = \begin{cases} -x \log x & \text{if } x > 0 \\ 0 & \text{if } x = 0, \end{cases} \tag{7.3}$$

and log denotes the logarithmic function based 2.

If $\mathbb{H}$ is a finite-dimensional Hilbert space, then all linear operators on $\mathbb{H}$ are finite ranks. Therefore, Definition 7.1.1 and Definition 8.1.3 can be treated as the definition of von Neumann entropy and quantum relative entropy for states on a finite-dimensional Hilbert space.

We now define the von Neumann entropy on infinite-dimensional Hilbert space $\mathbb{H}$ as follows.

Definition 7.1.2. The *von Neumann entropy* $H(\rho)$ of the quantum state $\rho \in \mathcal{S}(\mathbb{H})$ is defined by

$$H(\rho) = -\sum_i \langle i|\rho \log \rho|i\rangle_{\mathbb{H}}, \tag{7.4}$$

where $\{|i\rangle_{\mathbb{H}}\}_{i=1}^{+\infty} \subset \mathbb{H}$ is a basis of the eigenvectors of ρ.

Writing a quantum state $\rho \in \mathcal{S}(\mathbb{H})$ in the form of its spectral decomposition $\rho = \sum_i p_i |i\rangle_{\mathbb{H}}\langle i|$, $p_i > 0$ and $\sum_i p_i = 1$. Then the von Neumann entropy $H(\rho)$ defined in (7.4) can be expressed in the form that is comparable with Shannon entropy $S(X)$ (see (7.1) for a definition of Shannon entropy) as follows:

$$H(\rho) = -\sum_i p_i \log p_i = \sum_i \eta(p_i). \tag{7.5}$$

As noted in Wehrl [175], the von Neumann entropy $H(\cdot)$ expressed in the form of (7.5) provides a canonical approximation as follows.

If $\rho \in \mathcal{S}(\mathbb{H})$ if of finite rank, i. e., only finitely many eigenvalues of ρ are positive (say $p_i > 0$ for $i = 1, 2, \ldots, N$ and all other p_i are zero), then it is clear that $H(\rho) < +\infty$. Now let ρ be an arbitrary quantum state on $\mathbb{H}$. The canonical approximations of ρ can be written as

$$\rho^{(N)} = \frac{\sum_{i=1}^N p_i |i\rangle_{\mathbb{H}}\langle i|}{\sum_{i=1}^N p_i} \to \rho \quad \text{in } \|\cdot\|_1\text{-norm as } N \to +\infty.$$

Since

$$H(\rho^{(N)}) = -\sum_{i=1}^N \frac{p_i}{\sum_{i=1}^N p_i} \log\left(\frac{p_i}{\sum_{i=1}^N p_i}\right) = -\frac{\sum_{i=1}^N p_i \log p_i}{\sum_{i=1}^N p_i + \log(\sum_{i=1}^N p_i)},$$

we have $\lim_{N\to+\infty} H(\rho^{(N)}) = H(\rho)$ because

$$-\lim_{N\to+\infty} p_i \log p_i = H(\rho) \quad \text{and} \quad \lim_{N\to+\infty} \sum_{i=1}^N p_i = 1.$$

This fact may be of use in generalizing related theorems that hold for finite- dimensional cases to infinite-dimensional cases.

7.2 Properties of $H(\cdot)$ on $\mathcal{S}(\mathbb{H})$

The properties of $H(\cdot)$ along with their physical interpretation have been systematically explored in Wehrl [175].

We need the following two results, namely Klein's inequality (see Lindblad [107]) and the Ky Fan inequality [47], in order to establish important properties of $H(\cdot)$ on $\mathcal{S}(\mathbb{H})$.

Lemma 7.2.1 (Klein's inequality). *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex C^1 function and $\rho, \sigma \in \mathcal{S}(\mathbb{H})$. Then* $\operatorname{tr}[f(\rho) - f(\sigma) - (\rho - \sigma)f'(\sigma)] \geq 0$.

Proof. Let $\{\phi_i\}_{i=1}^{+\infty}$ and $\{\psi_i\}_{i=1}^{+\infty}$, respectively, be the eigenvectors of ρ and σ that form complete orthonormal bases of $\mathbb{H}$. Let $\{a_i\}_{i=1}^{+\infty}$ and $\{b_i\}_{i=1}^{+\infty}$ be the eigenvalues corresponding to $\{\phi_i\}_{i=1}^{+\infty}$ and $\{\psi_i\}_{i=1}^{+\infty}$, respectively. In this case, $f(\rho)\phi_i = f(a_i)\phi_i$ for $i = 1, 2, \ldots$. Then, by the definition of trace, we have

$$\begin{aligned}
&\operatorname{tr}[f(\rho) - f(\sigma) - (\rho - \sigma)f'(\sigma)] \\
&\quad = \sum_{i=1}^{+\infty} \langle \phi_i | f(\rho) - f(\sigma) - (\rho - \sigma)f'(\sigma) | \phi_i \rangle_{\mathbb{H}}
\end{aligned} \tag{7.6}$$

Now $\langle \phi_i | f(\rho) | \phi_i \rangle_{\mathbb{H}} = f(a_i)$ and $\langle \phi_i | \rho f'(\sigma) | \phi_i \rangle_{\mathbb{H}} = a_i \langle \phi_i | f'(\sigma) | \phi_i \rangle_{\mathbb{H}}$.

Each term in the sum of the summands (7.5) equals

$$f(a_i) - \langle \phi_i | f(\sigma) | \phi_i \rangle_{\mathbb{H}} - a_i \langle \phi_i | f'(\sigma) | \phi_i \rangle_{\mathbb{H}} + \langle \phi_i \sigma f'(\sigma) | \phi_i \rangle_{\mathbb{H}}. \tag{7.7}$$

Using the closure relation $\sum_{i=1}^{+\infty} |\psi_i\rangle_{\mathbb{H}} \langle \psi_i| = \mathbf{1}$, (7.5) can be rewritten as

$$\begin{aligned}
&\operatorname{tr}[f(\rho) - f(\sigma) - (\rho - \sigma)f'(\sigma)] \\
&\quad = \sum_j \left| \langle \phi_i, \psi_j \rangle_{\mathbb{H}} \right|^2 (f(a_i) - f(b_j) - (a_i - b_j)f'(b_j)).
\end{aligned}$$

Now, since $f : \mathbb{R} \to \mathbb{R}$ is convex, we have $f(a_i) - f(b_j) \geq (a_i - b_j)f'(b_j)$, which proves the lemma. □

Lemma 7.2.2 (Ky Fan inequality). *Let $\rho, \sigma \in \mathcal{S}(\mathbb{H})$ be quantum states with the sets of eigenvalues $\{p_k(\rho)\}_{k=1}^{+\infty}$ and $\{p_k(\sigma)\}_{k=1}^{+\infty}$, respectively, each arranged in decreasing order. Then, for each $k = 1, 2, \ldots$,*

$$p_1(\rho) + p_2(\rho) + \cdots + p_k(\rho) \leq \operatorname{tr}[|\rho - \sigma|] + p_1(\sigma) + \cdots + p_k(\sigma).$$

Proof. For $\rho, \sigma \in \mathcal{S}(\mathbb{H})$, we write the spectral decompositions of ρ and σ as

$$\rho = \sum_{i=1}^{+\infty} p_i(\rho) |\phi_i(\rho)\rangle_{\mathbb{H}} \langle \phi_i(\rho)| \quad \text{and} \quad \sigma = \sum_{i=1}^{+\infty} p_i(\sigma) |\phi_i(\sigma)\rangle_{\mathbb{H}} \langle \phi_i(\sigma)|,$$

respectively, where $(p_i(\rho))_{i=1}^{+\infty}$ (resp., $(p_i(\sigma))_{i=1}^{+\infty}$) is the sequence of eigenvalues of ρ (resp., σ) arranged in decreasing order and $(\phi_i(\rho))_{i=1}^{+\infty}$ (resp., $(\phi_i(\sigma))_{i=1}^{+\infty}$) be its corresponding unit eigenvectors that forms an orthonormal basis of $\mathbb{H}$. Without loss of generality, we can assume that $\rho \geq \sigma$ (and hence, $p_i(\rho) \geq p_i(\sigma)$), then for any $k = 1, 2, \ldots$ we have

$$\begin{aligned}
\operatorname{tr}[|\rho - \sigma|] &= \operatorname{tr}[\rho - \sigma] \\
&= \sum_{i=1}^{+\infty} (\langle \phi_i(\rho), \rho \phi_i(\rho) \rangle_{\mathbb{H}} - \langle \phi_i(\sigma), \sigma \phi_i(\sigma) \rangle_{\mathbb{H}})
\end{aligned}$$

$$= \sum_{i=1}^{+\infty} (\langle \phi_i(\rho), p_i(\rho)\phi_i(\rho)\rangle_{\mathbb{H}} - \langle \phi_i(\sigma), p_i(\sigma)\phi_i(\sigma)\rangle_{\mathbb{H}})$$
$$\geq \sum_{i=1}^{k} (p_i(\rho)\langle \phi_i(\rho), \phi_i(\rho)\rangle_{\mathbb{H}} - p_i(\sigma)\langle \phi_i(\sigma), \phi_i(\sigma)\rangle_{\mathbb{H}}) = \sum_{i=1}^{k} (p_i(\rho) - p_i(\sigma)).$$

This proves the lemma. □

Important properties of von Neumann entropy $H(\cdot) : \mathcal{S}(\mathbb{H}) \to [0, +\infty]$ are provided in the following propositions.

Proposition 7.2.3 (Positivity). *$H(\rho) \geq 0$, for all $\rho \in \mathcal{S}(\mathbb{H})$ and $H(\rho) = 0$ if and only if ρ is a pure state.*

Proof. Positivity of $H(\cdot) : \mathcal{S}(\mathbb{H}) \to [0, +\infty]$ follows from (7.5), since $H(\rho) = -\sum_i p_i \log p_i$ for $\rho = \sum_i p_i |i\rangle_{\mathbb{H}}\langle i|$, $p_i > 0$ and $\sum_i p_i = 1$, and the positive of the function $-x \log x$ for $0 \leq x \leq 1$. Note that $\{p_i\}_{i=1}^{+\infty}$, the eigenvalues of a pure state $|\psi\rangle_{\mathbb{H}}\langle\psi|$, consists of 1's or 0's. Therefore, $H(|\psi\rangle_{\mathbb{H}}\langle\psi|) = \sum_{i=1}^{+\infty} -p_i \log p_i = 0$. This proves the proposition. □

The von Neumann entropy is preserved under unitary transformations as shown in the following proposition.

Proposition 7.2.4 (Unitary invariance). *Let $\sigma = \mathbf{U}\rho\mathbf{U}^*$ for an unitary operator $\mathbf{U}$. Then $H(\sigma) = H(\rho)$.*

Proof. It is clear that any quantum state is unitary invariant because unitary transformation only changes the basis of the state. Since $H(\rho)$ depends only on the eigenvalues of ρ by Definition 7.1.2 and is independent of the basis and, therefore, independent of eigenvalues used in expressing the state. We, therefore, have $H(\rho) = H(\mathbf{U}\rho\mathbf{U}^*)$ for all $\rho \in \mathcal{S}(\mathbb{H})$. This proves the proposition. □

The above proposition shows that two unitarily equivalent quantum states have the same von Neumann entropy. However, the converse is not true in general. For a finite-dimensional Hilbert space, He, Hou and Li [64] give a sufficient and necessary condition of unitary equivalence of quantum states associated with the von Neumann entropy. It states that $H(\lambda\rho + (1-\lambda)\frac{\mathbf{I}}{n}) = H(\lambda\sigma + (1-\lambda)\frac{\mathbf{I}}{n}), \forall \lambda \in [0, 1]$ if and only if there exists a unitary $\mathbf{U}$ such that $\rho = \mathbf{U}\sigma\mathbf{U}^*$. The readers are referred to the above references for details.

Proposition 7.2.5 (Concavity). *The von Neumann entropy $H(\cdot) : \mathcal{S}(\mathbb{H}) \to [0, +\infty]$ is concave. That is, for all $\rho_1, \rho_2 \in \mathcal{S}(\mathbb{H})$ and $\lambda_1, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1$,*

$$H(\lambda_1\rho_1 + \lambda_2\rho_2) \geq \lambda_1 H(\rho_1) + \lambda_2 H(\rho_2). \tag{7.8}$$

Equation (7.8) holds for $\lambda_1, \lambda_2 > 0$ only if $\rho_1 = \rho_2$ or $H(\rho_1)$ or $H(\rho_2)$, respectively, equals $+\infty$.

Proof. Assume that $\rho = \lambda_1\rho_1 + \lambda_2\rho_2 = \sum_{k=1}^{+\infty} p_k|k\rangle_{\mathbb{H}}\langle k|$, where $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$ is the set of unit eigenvectors of ρ corresponding to the set of its eigenvalues $\{p_k\}_{k=1}^{+\infty}$ of ρ. Then, because of concavity of $\eta(x) = -x\log x$ (where $x > 0$),

$$\begin{aligned} H(\rho) &= -\sum_k p_k \log p_k = \sum_k \eta(\langle k|\rho|k\rangle_{\mathbb{H}}) \\ &\geq \lambda_1 \sum_k \eta(\langle k|\rho_1|k\rangle_{\mathbb{H}}) + \lambda_2 \sum_k \eta(\langle k|\rho_2|k\rangle_{\mathbb{H}}) \\ &\geq \lambda_1 \sum_k \langle k|\eta(\rho_1)|k\rangle_{\mathbb{H}} + \lambda_2 \sum_k \langle k|\eta(\rho_2)|k\rangle_{\mathbb{H}} = \lambda_1 H(\rho_1) + \lambda_2 H(\rho_2). \end{aligned}$$

This proves the concavity of the von Neumann entropy $H(\cdot)$. □

The proof of the following proposition will be provided in Corollary 8.1.7.

Proposition 7.2.6 (Generalized concavity). *Let ρ and σ be two quantum states on $\mathbb{H}$, then for all $\lambda \in [0,1]$,*

$$H(\lambda\rho + (1-\lambda)\sigma) \geq \lambda H(\rho) + (1-\lambda)H(\sigma) + \frac{\lambda(1-\lambda)}{2}\|\rho - \sigma\|_1. \tag{7.9}$$

As shown in the following result, $H(\cdot)$ is a continuous and bounded functions in finite dimensions. However, it can take the value $+\infty$ on a dense subset of $\mathcal{S}(\mathbb{H})$ when $\mathbb{H}$ is infinite-dimensional. As it is well known (see, e. g., Wehrl [175] and Shirokov [144]) that the properties of the entropy for infinite- and finite-dimensional Hilbert spaces differ quite substantially: in the latter case, the entropy is bounded continuous function on $\mathcal{S}(\mathbb{H})$, while in the former it is discontinuous (however, it is lower semicontinuous) at every point, and infinite "most everywhere" in the sense that the set of states with finite entropy is a first category subset of $\mathcal{S}(\mathbb{H})$, which we do not intend to address here.

Proposition 7.2.7 (Lower semicontinuity). *Let $(\rho_n)_{n=1}^{+\infty}$ be a sequence of states in $\mathcal{S}(\mathbb{H})$ such that $\lim_{n\to+\infty}\|\rho_n - \rho\|_1 = 0$ for some $\rho \in \mathcal{S}(\mathbb{H})$. Then*

$$H(\rho) \leq \liminf_{n\to+\infty} H(\rho_n).$$

Proof. Let $(\rho_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H})$ be a sequence of quantum states on $\mathbb{H}$ such that $\lim_{n\to+\infty}\|\rho_n - \rho\|_1 = 0$ for some $\rho \in \mathcal{S}(\mathbb{H})$. For each $n \in \mathbb{N}$, consider the following spectral decompositions of ρ_n and ρ, respectively,

$$\rho_n = \sum_{k=1}^{+\infty} p_k(\rho_n)|\phi_k(\rho_n)\rangle_{\mathbb{H}}\langle\phi_k(\rho_n)| \quad \text{and} \quad \rho = \sum_{k=1}^{+\infty} p_k(\rho)|\phi_k(\rho)\rangle_{\mathbb{H}}\langle\phi_k(\rho)|,$$

where $(\phi_k(\rho_n))_{k=1}^{+\infty}$ and $(\phi_k(\rho))_{k=1}^{+\infty}$ are eigenvectors of ρ_n and ρ, respectively, corresponding to eigenvalues $p_k(\rho_n) > 0$ and $p_k(\rho) > 0$ (arranged in decreasing order) By Ky Fan's inequality (see Lemma 7.2.2), we have

$$p_1(\rho_n) + p_2(\rho_n) + \cdots + p_k(\rho_n) \leq \operatorname{tr}[|\rho_n - \rho|] + p_1(\rho) + \cdots + p_k(\rho)$$

and also by writing $\rho_n = (\rho_n - \rho) + \rho$

$$p_1(\rho) + p_2(\rho) + \cdots + p_k(\rho) \leq \operatorname{tr}[|\rho_n - \rho|] + p_1(\rho_n) + p_2(\rho_n) + \cdots + p_k(\rho_n).$$

Therefore,

$$0 \leq \lim_{n\to+\infty} \left|(p_1(\rho_n) - p_1(\rho)) + \cdots + (p_k(\rho_n) - p_k(\rho))\right| \leq \lim_{n\to+\infty} \operatorname{tr}[|\rho_n - \rho|] = 0.$$

Eventually, $\lim_{n\to+\infty} p_k(\rho_n) = p_k(\rho)$ for each $k = 1, 2, \ldots$. Thus,

$$-\sum_{i=1}^{k} p_i(\rho) \log p_i(\rho) = -\lim_{n\to+\infty}\left(\sum_{i=1}^{k} p_i(\rho_n) \log p_i(\rho_n)\right)$$

and

$$\begin{aligned} H(\rho) &= \sup_k \left(\sum_{i=1}^{k} -p_i(\rho) \log p_i(\rho)\right) \\ &\leq \liminf_{n\to+\infty} \sup_k \left(\sum_{i=1}^{k} -p_i(\rho_n) \log p_i(\rho_n)\right) = \liminf_{n\to+\infty} H(\rho_n). \end{aligned}$$

This proves the lower semicontinuity. □

Proposition 7.2.8 (Subadditivity). *For the reduced states $\rho_1 = \operatorname{tr}_{\mathbb{H}_2}[\rho]$, $\rho_2 = \operatorname{tr}_{\mathbb{H}_1}[\rho]$ of $\rho \in \mathcal{S}(\mathbb{H}_{12})$ (where $\mathbb{H}_{12} = \mathbb{H}_1 \otimes \mathbb{H}_2$), $H(\rho) \leq H(\rho_1) + H(\rho_2)$ and the mapping $\rho_{12} \mapsto H(\rho_1) - H(\rho_{12})$ is convex on the set of positive, trace-class operators on $\mathbb{H}_{12}$.*

Proof. Writing $\rho = \rho_{12}$, it follows that

$$\begin{aligned} H(\rho) &= -\operatorname{tr}[\rho_{12} \log \rho_{12}] \\ &\leq -\operatorname{tr}[\rho_{12} \log(\rho_1 \otimes \rho_2)] \quad (\text{since } \log(\rho_1 \otimes \rho_2) \leq \log \rho_{12}) \\ &= -\operatorname{tr}[\rho_{12} \log(\rho_1 \otimes \mathbf{I}_2) + \rho_{12} \log(\mathbf{I}_1 \otimes \rho_2)] \\ &= -\operatorname{tr}[\rho_{12}(\log \rho_1 \otimes \mathbf{I}_2) + \rho_{12}(\mathbf{I}_1 \otimes \log \rho_2)] \\ &= -\operatorname{tr}[\rho_{12}(\log \rho_1 \otimes \mathbf{I}_2)] - \operatorname{tr}[\rho_{12}(\mathbf{I}_1 \otimes \log \rho_2)] \\ &= -\operatorname{tr}_1[\rho_1 \log \rho_1] - \operatorname{tr}_2[\rho_2 \log \rho_2] \quad (\text{by Proposition 2.8.2}) \\ &= H(\rho_1) + H(\rho_2), \end{aligned}$$

where $\mathbf{I}_i$ is the identity operator of $\mathbb{H}_i$ for $i = 1, 2$. This proves the subadditivity of $H(\cdot)$. □

We have the following additivity as a special case of the above proposition when $\rho_{12} = \rho_1 \otimes \rho_2$.

Corollary 7.2.9 (Additivity). *Let $\rho_i \in \mathbb{H}_i$ for $i = 1, 2$. Then $H(\rho_1 \otimes \rho_2) = H(\rho_1) + H(\rho_2)$.*

The following strong subadditivity of von Neumann entropy (SSA) has been known and well appreciated in classical probability theory and information theory. Its extension to quantum mechanical entropy (the von Neumann entropy) was conjectured by Robinson and Ruelle [130], Lanford and Robinson [103] and proved by Lieb and Ruskai [105]. It is a basic theorem in modern quantum information theory. SSA concerns the relation between the entropies of various subsystems of a larger system consisting of three subsystems (or of one system with three degrees of freedom). The proof of this relation in the classical case is quite easy but the quantum case is difficult because of the noncommutativity of the density matrices describing the subsystems.

The proof of the following proposition will be postponed until after the relative entropy is introduced in the next chapter.

Proposition 7.2.10 (Strong subadditivity). *Let $\mathbb{H}_{123} = \mathbb{H}_1 \otimes \mathbb{H}_2 \otimes \mathbb{H}_3$ and let $\rho_{123} \in \mathcal{S}(\mathbb{H}_{123})$. For $i, j, k = 1, 2, 3$, denote the reduced states $\mathrm{tr}_{\mathbb{H}_i \otimes \mathbb{H}_j}[\rho_{123}]$ by ρ_k and $\mathrm{tr}_{\mathbb{H}_k}[\rho_{123}]$ by ρ_{ij}. Then*

$$H(\rho_{123}) + H(\rho_2) \leq H(\rho_{12}) + H(\rho_{23}) \tag{7.10}$$

and

$$H(\rho_1) + H(\rho_2) \leq H(\rho_{13}) + H(\rho_{23}). \tag{7.11}$$

Proposition 7.2.11 (Dominated convergence theorem for von Neumann entropy). *Let ρ_n, ρ be elements in $\mathcal{S}(\mathbb{H})$ satisfying the following conditions:* (i) *$\rho_n \to \rho$ weakly as $n \to +\infty$;* (ii) *$\rho_n \leq \mathbf{A}$ for all n for some compact operator $\mathbf{A}$; and* (iii) *$-\sum_k a_k \log a_k < +\infty$ for the eigenvalues $\{a_k\}_{k=1}^{+\infty}$ of $\mathbf{A}$. Then $H(\rho_n) \to H(\rho)$ as $n \to +\infty$.*

Proof. For each compact operator $\mathbf{A}$ on $\mathbb{H}$ and each $\beta > 0$, it is easy to show that $F(\rho, \beta, \mathbf{A}) \equiv \mathrm{tr}[\rho \mathbf{A} - \beta^{-1} H(\rho)]$ is lower semicontinuous in ρ. It is clear that

$$\lim_{n \to +\infty} \mathrm{tr}[\rho_n \mathbf{A}] = \mathrm{tr}[\rho \mathbf{A}] \quad \text{if } \mathrm{tr}[e^{-\beta \mathbf{A}}] < +\infty \text{ for all } 0 < \beta < +\infty$$

Now $\lim_{n \to +\infty} \|\rho_n - \rho\|_1 = 0$ and $\lim_{n \to +\infty} \mathrm{tr}[\rho_n \mathbf{A}] = \mathrm{tr}[\rho \mathbf{A}]$ together implies that

$$H(\rho) = \liminf_{n \to +\infty} H(\rho_n) \quad \text{and} \quad -H(\rho) = \liminf_{n \to +\infty} (-H(\rho_n)) = -\limsup_{n \to +\infty} H(\rho_n)$$

This proves that $\lim_{n\to+\infty} H(\rho_n) = H(\rho)$ under the condition that $\mathrm{tr}[e^{-\beta\mathbf{A}}] < +\infty$ for all $0 < \beta < +\infty$, which is met since $\mathbf{A}$ is a compact operator on $\mathbb{H}$. This proves the lemma. □

7.3 $H(\cdot)$ on subsets of $\mathcal{S}(\mathbb{H})$

In this section, properties of von Neumann entropy on some proper subsets of $\mathcal{S}(\mathbb{H})$ will be systematically investigated. These proper subsets of $\mathcal{S}(\mathbb{H})$ include:
(i) Closed convex subset $\mathcal{A}$ of $\mathcal{S}(\mathbb{H})$ with $\sup_{\rho\in\mathcal{A}} H(\rho) < +\infty$;
(ii) Compact subset $\mathcal{K}_{\mathbf{H}}(h)$ of $\mathcal{S}(\mathbb{H})$, defined by

$$\mathcal{K}_{\mathbf{H}}(h) = \{\rho \in \mathcal{S}(\mathbb{H}) \mid \mathrm{tr}[\rho\mathbf{H}] \le h\},$$

where $\mathbf{H}$ is an $\mathfrak{H}$-operator and $h > 0$ (see (3.2) for definition of an $\mathfrak{H}$-operator); and
(iii) Arbitrary closed subset $\mathcal{K}$ of $\mathcal{S}(\mathbb{H})$.

The von Neumann entropies constrained on these proper subsets determine characterizations of capacities of constrained channels in later chapters.

The presentation of this section is largely based on results obtained by Holevo [78] and Shirokov [141, 144].

7.3.1 Closed convex subsets $\mathcal{A} \subset \mathcal{S}(\mathbb{H})$

Let $H(\cdot) : \mathcal{S}(\mathbb{H}) \to [0, +\infty]$ be the von Neumann entropy and let $\mathcal{A}$ be a convex closed subset of $\mathcal{S}(\mathbb{H})$ with $\sup_{\rho\in\mathcal{A}} H(\rho) < +\infty$. If this supremum is achieved at a particular state in $\mathcal{A}$, then this state is usually called the *Gibbs state* and denoted by $\Gamma(\mathcal{A})$.

Example 7.1. The following are three interesting examples of closed convex subsets $\mathcal{A} \subset \mathcal{S}(\mathbb{H})$:
(i) the set $\mathcal{A}$ consists of commuting states;
(ii) $\mathcal{A}$ is a convex hull of a finite collection of states; i. e., the smallest convex set containing the finite collection of states.
(iii) $\mathcal{A} = \Upsilon(\mathcal{B})$, where $\mathcal{B}$ is a compact subset of $\mathfrak{T}_+(\mathbb{K})$ with $\dim(\mathbb{K}) < +\infty$ and Υ is positive linear map from $\mathfrak{T}(\mathbb{K})$ to $\mathfrak{T}(\mathbb{H})$.

We have the following simple observation.

Lemma 7.3.1. *Let $\mathcal{A}$ be a closed convex subset of $\mathcal{S}(\mathbb{H})$ and let $(\rho_n)_{n=1}^{+\infty}$ be an arbitrary sequence of states in $\mathcal{A}$ such that*

$$\lim_{n\to+\infty} H(\rho_n) = \sup_{\rho\in\mathcal{A}} H(\rho) < +\infty.$$

Then this sequence converges to the uniquely defined state $\rho_*(\mathcal{A})$ in $\mathcal{A}$ under $\|\cdot\|_1$. If the Gibbs state $\Gamma(\mathcal{A})$ exists, then it coincides with the state $\rho_*(\mathcal{A})$ and the restriction of the entropy to the set $\mathcal{A}$ is continuous at the state $\Gamma(\mathcal{A})$.

Proof. Let $\epsilon > 0$ be an arbitrary positive number. By the assumption, there exists N_ϵ such that

$$H(\rho_n) > \sup_{\rho\in\mathcal{A}} H(\rho) - \epsilon, \quad \forall n \geq N_\epsilon.$$

The convexity of $H(\cdot)$ and (7.9) with $\lambda = \frac{1}{2}$ implies

$$\begin{aligned}\sup_{\rho\in\mathcal{A}} H(\rho) - \epsilon &\leq \frac{1}{2}H(\rho_{n_1}) + \frac{1}{2}H(\rho_{n_2})\\ &\leq H\left(\frac{1}{2}\rho_{n_1} + \frac{1}{2}\rho_{n_2}\right) - \frac{1}{8}\|\rho_{n_2} - \rho_{n_1}\|_1^2 \leq \sup_{\rho\in\mathcal{A}} H(\rho) - \frac{1}{8}\|\rho_{n_2} - \rho_{n_1}\|_1^2,\end{aligned}$$

and hence, $\|\rho_{n_2} - \rho_{n_1}\|_1 < \sqrt{8\epsilon}$ for all $n_1 \geq N_\epsilon$ and $n_2 \geq N_\epsilon$. Thus, the sequence $(\rho_n)_{n=1}^{+\infty}$ is a Cauchy sequence under the trace norm $\|\cdot\|_1$, and hence, it converges to a particular state ρ_* in $\mathcal{A}$. It is easy to see that this state ρ_* does not depend on the choice of the sequence $(\rho_n)_{n=1}^{+\infty}$, so it is determined only by the set $\mathcal{A}$. Denote this state by $\rho_*(\mathcal{A})$. If the Gibbs state $\Gamma(\mathcal{A})$ exists, then by the above observation it coincides with the state $\rho_*(\mathcal{A})$. The continuity assertion follows from lower semicontinuity of the entropy H. This proves the lemma. □

7.3.2 $\mathcal{K}_{\mathbf{H}}(h)$

Let $\mathbf{H}$ be an $\mathfrak{H}$-operator on $\mathbb{H}$. That is, $\mathbf{H}$ is an unbounded positive linear operator on $\mathbb{H}$ with discrete eigenvalues and with finite multiplicity

$$0 \leq h_m(\mathbf{H}) < h_{d+1} \leq h_{d+2} \leq \cdots,$$

where $h_m(\mathbf{H})$ is the lowest eigenvalue of $\mathbf{H}$ with d multiplicity for some $1 \leq d < +\infty$. That is,

$$h_m(\mathbf{H}) = h_1 = h_2 = \cdots = h_d < h_{d+1} \leq \cdots.$$

Let $\mathbb{H}_m(\mathbf{H}) \subset \mathbb{H}$ be the eigenspace corresponding to $h_m(\mathbf{H})$. In this case, $\mathbb{H}_m(\mathbf{H})$ is a d-dimensional subspace of $\mathbb{H}$.

Let $\mathbf{Q}_n$ be the spectral projector of the operator $\mathbf{H}$ corresponding to the lowest n eigenvalues $h_m(\mathbf{H}) = h_1 = h_2 = \cdots = h_d < h_{d+1} \leq \cdots \leq h_n$. According to (3.2), we define $\mathrm{tr}[\rho\mathbf{H}]$ by

$$\mathrm{tr}[\rho\mathbf{H}] = \lim_{n\to\infty} \mathrm{tr}[\rho\mathbf{Q}_n\mathbf{H}],$$

where the sequence on the right-hand side above is monotonely nondecreasing.

Recall from (3.3) that $\mathcal{K}_{\mathbf{H}}(h) \subset \mathcal{S}(\mathbb{H})$ is defined by

$$\mathcal{K}_{\mathbf{H}}(h) = \{\rho \in \mathcal{S}(\mathbb{H}) \mid \mathrm{tr}[\rho\mathbf{H}] \le h\}.$$

We recall the following simple observations on the set $\mathcal{K}_{\mathbf{H}}(h)$:

1. If $\mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{H})$, then $\mathcal{K} \subset \mathcal{K}_{\mathbf{H}}(h)$ for some $\mathfrak{H}$-operator $\mathbf{H}$ and some $h > 0$ (from Theorem 3.2.5);
2. $\overline{\bigcup_{h\in\mathbb{R}_+} \mathcal{K}_{\mathbf{H}}(h)}^{\|\cdot\|_1} = \mathcal{S}(\mathbb{H})$, where the closure is taken under the $\|\cdot\|_1$-norm. It is clear that $\overline{\bigcup_{h\in\mathbb{R}_+} \mathcal{K}_{\mathbf{H}}(h)}^{\|\cdot\|_1} \subseteq \mathcal{S}(\mathbb{H})$. On the other hand, if $\rho \in \mathcal{S}(\mathbb{H})$, then for every $\epsilon > 0$ there exists $h = h(\epsilon,\rho) > 0$ such that $\rho \in \mathcal{K}_{\mathbf{H}}(h)$. Therefore, $\mathcal{S}(\mathbb{H}) \subseteq \overline{\bigcup_{h\in\mathbb{R}_+} \mathcal{K}_{\mathbf{H}}(h)}^{\|\cdot\|_1}$. Consequently, $\overline{\bigcup_{h\in\mathbb{R}_+} \mathcal{K}_{\mathbf{H}}(h)}^{\|\cdot\|_1} = \mathcal{S}(\mathbb{H})$.

We follow Shirokov [141] for the definitions of an increasing coefficient of the $\mathfrak{H}$-operator $\mathbf{H}$ and decreasing coefficient of $\sigma \in \mathcal{S}(\mathbb{H})$ given below:

$$\lambda^{\dagger}(\mathbf{H}) = \begin{cases} \inf\{\lambda > 0 \mid \mathrm{tr}[e^{-\lambda\mathbf{H}}] < +\infty\} & \text{if } \mathrm{tr}[e^{-\lambda\mathbf{H}}] < +\infty \text{ for some } \lambda > 0, \\ +\infty, & \text{if } \mathrm{tr}[-\lambda\mathbf{H}] = +\infty \text{ for all } \lambda > 0. \end{cases}$$

The real number $\lambda^{\dagger}(\mathbf{H})$ defined above is called the *increasing coefficient* of the $\mathfrak{H}$-operator $\mathbf{H}$.

Let σ be an arbitrary state in $\mathcal{S}(\mathbb{H})$. In what follows, we use the *decreasing coefficient* $\lambda_{\dagger}(\sigma)$ of the state σ defined as

$$\lambda_{\dagger}(\sigma) = \inf\{\lambda > 0 \mid \mathrm{tr}[\lambda\sigma] < +\infty\} \in [0,1]. \tag{7.12}$$

An example of an increasing coefficient and decreasing coefficient is given below.

Example 7.2. Let σ is a full rank state. Then $\mathbf{H} = -\log\sigma$ is an $\mathfrak{H}$-operator and $\lambda_{\dagger}(\sigma) = \lambda^{\dagger}(\mathbf{H})$. It is easy to see that $\lambda_{\dagger}(\sigma) < 1$ implies finiteness of the von Neumann entropy $H(\sigma)$ but there exist states σ with finite entropy (such as the states with spectrum $\{a((k+1)\log^3(k+1))^{-1}\}$) such that $\lambda_{\dagger}(\sigma) = 1$. The role of these states is shown in Proposition 7.3.9.

While it is known from Theorem 3.2.5 that for every compact subset $\mathcal{K}$ of $\mathcal{S}(\mathbb{H})$, there exists an $\mathfrak{H}$-operator $\mathbf{H}$ and an $h > 0$ such that $\mathcal{K} \subset \mathcal{K}_{\mathbf{H}}(h)$, the similar statement can be made for some specific closed convex subsets $\mathcal{A}$ such as those stated in Example 7.1 of $\mathcal{S}(\mathbb{H})$.

The following proposition follows from Example 7.2 since $\mathcal{A}$ that consists of commutative states is a compact subset of $\mathcal{S}(\mathbb{H})$.

Proposition 7.3.2. *Let $\mathcal{A}$ be the closed convex subset of $\mathcal{S}(\mathbb{H})$ that consists of all commutative states. Then there exist an $\mathfrak{H}$-operator $\mathbf{H}$ and an $h > 0$ such that $\mathcal{A} \subset \mathcal{K}_{\mathbf{H}}(h)$ if $\lambda^{\dagger}(\mathbf{H}) < +\infty$.*

In the following, we use the concept of relative entropy for finite- and infinite-dimensional Hilbert space $\mathbb{H}$. These are briefly defined below but their detailed properties are the subjects of investigation in the next chapter.

Let $\mathbf{A}$ and $\mathbf{B}$ be two positive finite rank operators on the Hilbert space $\mathbb{H}$. We define the relative entropy $H(\mathbf{A}\|\mathbf{B})$ of $\mathbf{A}$ and $\mathbf{B}$ as

$$H(\mathbf{A}\|\mathbf{B}) = \begin{cases} \operatorname{tr}[\mathbf{A}\log\mathbf{A} - \mathbf{A}\log\mathbf{B} + \mathbf{B} - \mathbf{A}], & \text{if } \operatorname{range}(\mathbf{A}) \subset \operatorname{range}(\mathbf{B}) \\ +\infty, & \text{otherwise.} \end{cases} \tag{7.13}$$

If ρ and σ are quantum states in $\mathcal{S}(\mathbb{H})$, the *relative entropy* $H(\rho\|\sigma)$ of states ρ and σ is defined by

$$H(\rho\|\sigma) = \begin{cases} \sum_i \langle i|\rho\log\rho - \rho\log\sigma\rangle_{\mathbb{H}}, & \text{if } \ker^{\perp}(\rho) \subset \ker^{\perp}(\sigma) \\ +\infty, & \text{otherwise,} \end{cases} \tag{7.14}$$

where $\{|i\rangle_{\mathbb{H}}\}_{i=1}^{+\infty} \subset \mathbb{H}$ is a basis of the eigenvectors of ρ.

Comparing the above two definitions (Definitions 8.1.3 and 8.1.4), we note that if $\mathbf{A}$ and $\mathbf{B}$ are quantum states on $\mathbb{H}$, then $\operatorname{tr}[\mathbf{B}] = \operatorname{tr}[\mathbf{A}] = 1$. Consequently, the two formulae (8.6) and (8.7) coincide.

Proposition 7.3.3. *Let $\mathcal{A}$ be a convex hull of finite subset of states in $\mathcal{S}(\mathbb{H})$. If the von Neumann entropy $H(\cdot)$ is continuous on $\mathcal{A}$, then there exists an $\mathfrak{H}$-operator $\mathbf{H}$ and an $h > 0$ such that $\mathcal{A} \subset \mathcal{K}_{\mathbf{H}}(h)$.*

Proof. If $\mathcal{A} = \operatorname{co}(\{\rho_i\}_{i=1}^n)$ (i. e., $\mathcal{A}$ is the convex hull of the finite set of states $\{\rho_i\}_{i=1}^n$), then the relative entropy $H(\rho_i\|\bar{\rho}) < +\infty$, where $\bar{\rho} = \frac{1}{n}\sum_{i=1}^n \rho_i$. Thus, validity of the assertion in the proposition in this case follows from the implication (2) $\Rightarrow$ (1) in Theorem 3.2.5 (with $\mathcal{A} = \{\rho_i\}_{i=1}^n$ and $\sigma = \bar{\rho}$). This proves the proposition. □

Proposition 7.3.4. *Let $\mathcal{A} = \Upsilon(\mathcal{B})$, where $\Upsilon : \mathfrak{B}(\mathbb{K}) \to \mathfrak{B}(\mathbb{H})$ and $\mathcal{B}$ is a compact subset of $\mathfrak{T}_+(\mathbb{K})$ with $\dim(\mathbb{K}) < +\infty$. If the von Neumann entropy $H(\cdot)$ is continuous on $\mathcal{A}$, then there exists an $\mathfrak{H}$-operator $\mathbf{H}$ and an $h > 0$ such that $\mathcal{A} \subset \mathcal{K}_{\mathbf{H}}(h)$.*

Proof. We may assume that the set $\mathcal{B}$ contains a full rank operator $\mathbf{B}_0$. Since the function $H(\cdot)$ is continuous on the closed convex set $\mathcal{A}$ it is bounded on this set, and hence, Theorem 3.2.5 implies existence of a positive operator $\mathbf{H}$ on $\mathbb{H}$ such that $\lambda^{\dagger}(\mathbf{H}) < +\infty$ and $\operatorname{tr}[\mathbf{H}\rho] \le h$ for all $\rho \in \mathcal{A}$ and some $h > 0$. Finiteness of $\lambda^{\dagger}(\mathbf{H})$ shows that $\mathbf{H} = \sum_{i=1}^{+\infty} h_i |i\rangle_{\mathbb{H}}\langle i|$, where $\{|i\rangle_{\mathbb{H}}\}_{i=1}^{+\infty}$ is an orthonormal basis of $\mathbb{H}$. Since $\operatorname{tr}[\mathbf{H}\Upsilon(\mathbf{B}_0)] \le h$ and $\mathbf{B}_0 \ge \lambda\mathbf{I}_{\mathbb{K}}$ for some $\lambda > 0$, we have

$$\operatorname{tr}[\mathbf{H}\Upsilon(\mathbf{I}_{\mathbb{K}})] = \sum_{i=1}^{+\infty} h_i \langle i|\Upsilon(\mathbf{I}_{\mathbb{K}})|i\rangle_{\mathbb{H}} = \operatorname{tr}\left[\sum_{i=1}^{+\infty} h_i \Upsilon^*(|i\rangle_{\mathbb{H}}\langle i|)\right] < +\infty,$$

and hence, the linear operator in the square bracket lies in $\mathfrak{B}(\mathbb{K})$. Thus, the function

$$\mathbf{B} \mapsto \operatorname{tr}[\mathbf{H}\Upsilon(\mathbf{B})] = \sum_{i=1}^{+\infty} h_i \langle i|\Upsilon(\mathbf{B})|i\rangle_{\mathbb{H}} = \operatorname{tr}\left[\sum_{i=1}^{+\infty} h_i \Upsilon^*(|i\rangle_{\mathbb{H}}\langle i|)\right]\mathbf{B}$$

is continuous on the compact set $\mathcal{B}$. For arbitrary compact subset $\mathcal{B}$ of $\mathfrak{T}_+(\mathbb{K})$, Dini's lemma (which states that a monotonically increasing sequence of continuous real-valued functions on a compact set, which converges pointwise to a continuous function then the convergence is uniform) implies uniform convergence of the series $\sum_{i=1}^{+\infty} h_i \Upsilon^*(|i\rangle_{\mathbb{H}}\langle i|)$ on $\mathcal{B}$, and hence, existence of a nondecreasing sequence $(y_i^{\mathcal{B}})_{i=1}^{+\infty}$ of positive numbers converging to $+\infty$ such that

$$\sup_{\mathbf{B}\in\mathfrak{T}_+(\mathbb{K})} \sum_{i=1}^{+\infty} y_i^{\mathcal{B}} h_i \Upsilon^*(|i\rangle_{\mathbb{H}}\langle i|) < +\infty$$

Let $\mathbf{H}^{\mathcal{B}} = \sum_{i=1}^{+\infty} y_i^{\mathcal{B}} h_i |i\rangle\langle i|$ be an $\mathfrak{H}$-operator such that $\lambda^{\dagger}(\mathbf{H}^{\mathcal{B}}) = 0$. Thus, we have

$$\sup_{\mathbf{B}\in\mathcal{B}} \operatorname{tr}[\mathbf{H}^{\mathcal{B}}\Upsilon(\mathbf{B})] = \sup_{\mathbf{B}\in\mathcal{B}} \operatorname{tr}\left[\sum_{i=1}^{+\infty} y_i^{\mathcal{B}} h_i \Upsilon^*(|i\rangle\langle i|)\right]\mathbf{B} < +\infty.$$

This proves the proposition. □

Lemma 7.3.5. *If σ is a quantum state with $\lambda_{\dagger}(\sigma) < 1$, then for arbitrary state ρ such that $H(\rho\|\sigma) < +\infty$ the von Neumann entropy $H(\rho)$ is finite and for all $\lambda > \lambda_{\dagger}(\sigma)$, the following identity holds:*

$$H(\rho\|(\operatorname{tr}[\sigma^{\lambda}])^{-1}\sigma^{\lambda}) = \lambda H(\rho\|\sigma) + \log(\operatorname{tr}[\sigma^{\lambda}]) - (1-\lambda)H(\rho).$$

If $\operatorname{tr}[\sigma^{\lambda_{\dagger}(\sigma)}] < +\infty$, *then above identity holds for* $\lambda = \lambda_{\dagger}(\sigma)$.

Proof. Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be an increasing sequence of spectral projectors of the state σ. Let $\mathbf{A}_n = \mathbf{P}_n\rho\mathbf{P}_n$ and $\mathbf{B}_n = \mathbf{P}_n\sigma\mathbf{P}_n$ be positive trace-class operators. By definition, we have

$$\begin{aligned}
H(\mathbf{A}_n\|\mathbf{B}_n^{\lambda}) &= \operatorname{tr}[\mathbf{A}_n\log\mathbf{A}_n - \mathbf{A}_n\log\mathbf{B}_n^{\lambda} + \mathbf{B}_n^{\lambda} - \mathbf{A}_n] \\
&= \operatorname{tr}[(\lambda + (1-\lambda))\mathbf{A}_n\log\mathbf{A}_n - \lambda\mathbf{A}_n\log\mathbf{B}_n + \mathbf{B}_n^{\lambda} - \mathbf{A}_n] \\
&= \lambda H(\mathbf{A}_n\|\mathbf{B}_n) + \operatorname{tr}[\mathbf{B}_n^{\lambda}] - \lambda\operatorname{tr}[\mathbf{B}_n] \\
&\quad - (1-\lambda)\operatorname{tr}[\mathbf{A}_n] - (1-\lambda)\operatorname{tr}[\mathbf{A}_n(-\log\mathbf{A}_n)].
\end{aligned}$$

Since $\mathbf{B}_n^{\lambda} = \mathbf{P}_n\sigma^{\lambda}\mathbf{P}_n$, Lindblad's results (see also Lemma 7.2.7) imply

$$\lim_{n\to+\infty}\operatorname{tr}[\mathbf{A}_n(-\log\mathbf{A}_n)] = H(\rho) \quad \text{and} \quad \lim_{n\to+\infty} H(\mathbf{A}_n\|\mathbf{B}_n^{\lambda}) = H(\rho\|\sigma^{\lambda})$$

for all $\lambda > \lambda_{\dagger}(\sigma)$. So, passing to the limit in the above equality, we obtain

$$H(\rho\|\sigma^{\lambda}) = \lambda H(\rho\|\sigma) + \mathrm{tr}[\sigma^{\lambda}] - 1 - (1-\lambda)H(\rho).$$

Thus, finiteness of $H(\rho\|\sigma)$ implies finiteness of $H(\rho)$ and of $H(\rho\|\sigma^{\lambda})$ for all $\lambda > \lambda_{\dagger}(\sigma)$. By noting that

$$H(\rho\|(\mathrm{tr}[\sigma^{\lambda}])^{-1}\sigma^{\lambda}) = H(\rho\|\sigma^{\lambda}) + \log(\mathrm{tr}[\sigma^{\lambda}]) - \mathrm{tr}[\sigma^{\lambda}] + 1,$$

we obtain the identity of the lemma. □

The following result, due originally to Shirokov [141], specifies conditions on $\lambda^{\dagger}(\mathbf{H})$ at which the von Neumann $H(\cdot)$ is either bounded or continuous on the set $\mathcal{K}_{\mathbf{H}}(h)$.

Proposition 7.3.6. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on the Hilbert space* $\mathbb{H}$ *and* h *be a positive number such that* $h > h_m(\mathbf{H})$. *Then:*
1. *The von Neumann entropy* $H(\cdot)$ *is bounded on the set* $\mathcal{K}_{\mathbf{H}}(h)$ *if and only if* $\lambda^{\dagger}(\mathbf{H}) < +\infty$.
2. *The von Neumann entropy* $H(\cdot)$ *is continuous on the set* $\mathcal{K}_{\mathbf{H}}(h)$ *if and only if* $\lambda^{\dagger}(\mathbf{H}) = 0$.

Proof. We follow the proof given in [141] below. Throughout this proof, we will assume that the $\mathfrak{H}$-operator takes the form $\mathbf{H} = \sum_{k=1}^{+\infty} h_k|k\rangle_{\mathbb{H}}\langle k|$, where $(h_k)_{k=1}^{+\infty}$ is a nondecreasing sequence of eigenvalues of $\mathbf{H}$ converging to infinity and $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty} \subset \mathbb{H}$ is an orthonormal basis of corresponding eigenvectors of $\mathbf{H}$. That is, $\mathbf{H}|k\rangle_{\mathbb{H}} = h_k|k\rangle_{\mathbb{H}}$ for $k = 1, 2, \ldots$, and $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$ form an orthonormal basis of $\mathbb{H}$. Let $d = \dim(\mathbb{H}_m(\mathbf{H}))$, so that $h_k = h_m(\mathbf{H})$ for $k = 1, 2, \ldots d$, and $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{d}$ is a basis of $\mathbb{H}_m(\mathbf{H})$. We begin with the proof of the first part of the proposition.

(1). ($\Rightarrow$) Suppose $\lambda^{\dagger}(\mathbf{H}) := \inf\{\lambda > 0 \mid \mathrm{tr}[\exp(-\lambda\mathbf{H})] < +\infty\} < +\infty$. Then there exists $\lambda > 0$ such that

$$\sigma = \frac{\exp(-\lambda\mathbf{H})}{\mathrm{tr}[\exp(-\lambda\mathbf{H})]}$$

is a state. By using nonnegativity of relative entropy and the definition of the set $\mathcal{K}_{\mathbf{H}}(h)$, we obtain

$$\begin{aligned} H(\rho) &= \lambda\,\mathrm{tr}[\rho\mathbf{H}] + \log(\mathrm{tr}[\exp(-\lambda\mathbf{H})]) - H(\rho\|\sigma) \\ &\leq \lambda h + \log(\mathrm{tr}[\exp(-\lambda\mathbf{H})]) \\ &\quad (\text{because } h > h_m(\mathbf{H}) \text{ and } H(\rho\|\sigma) \geq 0) < +\infty, \quad \forall \rho \in \mathcal{K}_{\mathbf{H}}(h), \end{aligned}$$

which means boundedness of $H(\rho)$ on $\mathcal{K}_{\mathbf{H}}(h)$.

($\Leftarrow$) On the other hand, suppose $\sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) < +\infty$. We first show that the equation

$$\sum_{k=1}^{n} h_k \exp(-\lambda h_k) = \sum_{k=1}^{n} h \exp(-\lambda h_k) \tag{7.15}$$

has the unique positive solution λ_n for all sufficiently large n and that the sequence $(\lambda_n)_{n=1}^{+\infty}$ is increasing. Note that equation (7.15) is equivalent to the equation $f_n(\lambda) = 0$, where

$$f_n(\lambda) = \sum_{k=1}^{n} (h_k - h) \exp(-\lambda(h_k - h)) = \exp(\lambda h) \sum_{k=1}^{n} (h_k - h) \exp(-\lambda h_k).$$

Since $f_n'(\lambda) = -\sum_{k=1}^{n}(h_k - h)^2 \exp(-\lambda(h_k - h)) < 0$, the function $f_n(\lambda)$ is strictly decreasing on $[0, +\infty[$ for each n. It is easy to see that

$$f_n(0) = \sum_{k=1}^{n} h_k - nh \quad \text{and} \quad \lim_{\lambda \to +\infty} f_n(\lambda) = -\infty \text{ if } h > h_m(\mathbf{H}).$$

Since the sequence $(h_k)_{k=1}^{+\infty}$ is nondecreasing and unbounded, $\sum_{k=1}^{n} h_k > nh$ (hence, $f_n(0) > 0$) for all sufficiently large n and the above observation imply existence of the unique positive solution λ_n of the equation $f_n(\lambda) = 0$ for each n. To show that $\lambda_{n+1} > \lambda_n$, it is sufficient to note that $f_{n+1}(\lambda) > f_n(\lambda)$ for all λ in $[0, +\infty[$ and for all n such that $h_n > h$.

For each sufficiently large n, consider the state

$$\rho_n = \left(\sum_{k=1}^{n} \exp(-\lambda_n h_k) \right)^{-1} \sum_{k=1}^{n} \exp(-\lambda_n h_k) |k\rangle_{\mathbb{H}} \langle k| \tag{7.16}$$

in $\mathcal{K}_{\mathbf{H}}(h)$.

This state is the maximum point of the entropy $H(\rho)$ on the subset $\mathcal{K}_{\mathbf{H}}^{(n)}(h)$ of $\mathcal{K}_{\mathbf{H}}(h)$ consisting of states supported by the linear hull of the vectors $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{n}$. Indeed, by using nonnegativity of the relative entropy and definition of the state ρ_n it is easy to see that

$$\begin{aligned} H(\rho) &= \lambda_n \operatorname{tr}[\rho \mathbf{H}] + \log\left(\sum_{k=1}^{n} \exp(-\lambda_n h_k) \right) - H(\rho \| \rho_n) \\ &\leq \lambda_n h + \log\left(\sum_{k=1}^{n} \exp(-\lambda_n h_k) \right) \\ &\quad (\text{because } h \geq \operatorname{tr}[\rho \mathbf{H}] \text{ and } H(\rho \| \rho_n) \geq 0) \end{aligned}$$

for all $\rho \in \mathcal{K}_{\mathbf{H}}^{(n)}(h)$ and that the equality of the inequality above takes place if and only if $\rho = \rho_n$. By using this and monotonicity of the logarithmic function, we obtain

$$H(\rho_n) = \lambda_n h + \log\left(\sum_{k=1}^{n} \exp(-\lambda_n h_k) \right) \geq \lambda_n (h - h_m(\mathbf{H})) \tag{7.17}$$

Since $h > h_m(\mathbf{H})$, the assumption $\sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) < +\infty$ implies boundedness of the sequence $(\lambda_n)_{n=1}^{+\infty}$. By this and due to the mentioned above monotonicity of this sequence, we conclude that there exists $\lim_{n\to+\infty}\lambda_n = \lambda^* < +\infty$. Since $\lambda_n \le \lambda^*$ for all n, the first equality in (7.17) implies

$$\sum_{k=1}^{n}\exp(-\lambda^* h_k) \le \sum_{k=1}^{n}\exp(-\lambda_k h_k) < \exp\Big(\sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho)\Big) < +\infty, \tag{7.18}$$

for all n, and hence,

$$\sum_{k=1}^{+\infty}\exp(-\lambda^* h_k) < +\infty. \tag{7.19}$$

This shows that $\lambda^\dagger(\mathbf{H}) \le \lambda^* < +\infty$. We conclude that the von Neumann entropy $H(\cdot)$ is bounded on $\mathcal{K}_{\mathbf{H}}(h)$ if and only if $\lambda^\dagger(\mathbf{H}) < +\infty$.

(2). For $n = 1, 2, \dots$, let $\mathcal{K}_{\mathbf{H}}^{(n)}(h)$ be as defined in the proof of part (1). Since $\mathcal{K}_{\mathbf{H}}(h) = \overline{\bigcup_n \mathcal{K}_{\mathbf{H}}^{(n)}(h)}$ and $\sup_{\rho\in\mathcal{K}_{\mathbf{H}}^{(n)}(h)} H(\rho) = H(\rho_n)$, lower semicontinuity of $H(\rho)$ implies that

$$\sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) = \lim_{n\to+\infty}\Big(\sup_{\rho\in\mathcal{K}_{\mathbf{H}}^{(n)}(h)} H(\rho)\Big) = \lim_{n\to+\infty} H(\rho_n).$$

By Lemma 7.3.1, the sequence of states $(\rho_n)_{n=1}^{+\infty}$ converges to the state $\rho_*(\mathcal{K}_{\mathbf{H}}(h))$. For each $n \in \mathbb{N}$, let $\mathbf{A}_n = \sum_{k=1}^{n}\exp(-\lambda_n h_k)|k\rangle_{\mathbb{H}}\langle k|$. Since $\lim_{n\to+\infty}\lambda_n = \lambda_*$, the sequence of operators $(\mathbf{A}_n)_{n=1}^{+\infty}$ converges to $\mathbf{A}_* = \sum_{k=1}^{+\infty}\exp(-\lambda^* h_k)|k\rangle_{\mathbb{H}}\langle k|$ in $\mathfrak{T}(\mathbb{H})$ in weak operator topology. Combining these observations, it is easy to see that

$$\lim_{n\to+\infty}\operatorname{tr}[\mathbf{A}_n] = \lim_{n\to+\infty}\sum_{k=1}^{n}\exp(-\lambda_n h_k) = \sum_{k=1}^{+\infty}\exp(-\lambda^* h_k) = \operatorname{tr}[\mathbf{A}_*], \tag{7.20}$$

since $\operatorname{tr}[|k\rangle_{\mathbb{H}}\langle k|] = 1$ for each k, and that

$$\rho_*(\mathcal{K}_{\mathbf{H}}(h)) = \lim_{n\to+\infty}\rho_n = \left(\sum_{k=1}^{+\infty}\exp(-\lambda^* h_k)\right)^{-1}\sum_{k=1}^{+\infty}\exp(-\lambda^* h_k)|k\rangle_{\mathbb{H}}\langle k|. \tag{7.21}$$

By using (7.17) and (7.20), we obtain

$$\sup_{\rho\in\mathcal{K}_{\mathbf{H},h}} H(\rho) = \lim_{n\to+\infty} H(\rho_n) = h\lambda^* + \log\left(\sum_{k=1}^{+\infty}\exp(-\lambda^* h_k)\right). \tag{7.22}$$

Lower semicontinuity of the entropy implies

$$H(\rho_*(\mathcal{K}_{\mathbf{H}}(h))) = \lambda^* \frac{\sum_{k=1}^{+\infty} h_k \exp(-\lambda^* h_k)}{\sum_{k=1}^{+\infty} \exp(-\lambda^* h_k)} + \log\left(\sum_{k=1}^{+\infty} \exp(-\lambda^* h_k)\right) \leq \lim_{n\to+\infty} H(\rho_n).$$

It follows from (7.22) that this inequality is equivalent to the inequality

$$\sum_{k=1}^{+\infty} h_k \exp(-\lambda^* h_k) \leq h \sum_{k=1}^{+\infty} \exp(-\lambda^* h_k). \tag{7.23}$$

Note that equality in this inequality implies that $\rho_*(\mathcal{K}_{\mathbf{H}}(h))$ is the Gibbs state $\Gamma(\mathcal{K}_{\mathbf{H}}(h))$. Conversely, by Lemma 7.3.1 if the Gibbs state $\Gamma(\mathcal{K}_{\mathbf{H}}(h))$ exists, then it coincides with $\rho_*(\mathcal{K}_{\mathbf{H}}(h))$, and hence, equality holds in (7.23). Thus, existence of the Gibbs state $\Gamma(\mathcal{K}_{\mathbf{H}}(h))$ is equivalent to equality in (7.23). So, to complete the proof of this part of the proposition it is sufficient to show that the inequality $h \leq h_*(\mathbf{H})$ is equivalent to equality in (7.23).

We first show that $\lambda^* > \lambda^\dagger(\mathbf{H})$ implies equality in (7.23). Consider the function

$$f(\lambda) = \lim_{n\to+\infty} f_n(\lambda) = \sum_{k=1}^{+\infty} (h_k - h)\exp(-\lambda(h_k - h)).$$

Since the series $\sum_{k=1}^{+\infty} h_k^p \exp(-\lambda h_k)$ converges uniformly on the interval $[\lambda^\dagger(\mathbf{H}) + \epsilon, +\infty[$ for arbitrary $p \in \mathbb{N}$ and $\epsilon > 0$, the function $f(\lambda)$ has a continuous derivative $f'(\lambda) = -\sum_{k=1}^{+\infty}(h_k - h)^2 \exp(-\lambda(h_k - h)) < 0$ on the interval $]\lambda^\dagger(\mathbf{H}), +\infty[$. By the construction, $f(\lambda_n) \geq f_n(\lambda_n) = 0$ for all sufficiently large n. This and continuity of the function $f(\lambda)$ at the point $\lambda^* \in]\lambda^\dagger(\mathbf{H}), +\infty[$ imply $f(\lambda^*) \geq 0$. Since (7.23) implies the converse inequality, we obtain $f(\lambda^*) = 0$, which means equality in (7.23).

If $h < h_*(\mathbf{H})$, then (finite or infinite) $f(\lambda^\dagger(\mathbf{H})) > 0$. Since (7.23) implies $f(\lambda^*) \leq 0$, this means $\lambda^* > \lambda^\dagger(\mathbf{H})$ and by the above observation $f(\lambda^*) = 0$. If $h = h_*(\mathbf{H})$, then $f(\lambda^\dagger(\mathbf{H})) = 0$, and hence, $\lambda^* = \lambda^\dagger(\mathbf{H})$. Indeed, if $\lambda^* > \lambda^\dagger(\mathbf{H})$ then by the above observation $f(\lambda^*) = 0 = f(\lambda^\dagger(\mathbf{H}))$ contradicting to the strict decreasing property of the function $f(\lambda)$.

If $h > h_*(\mathbf{H})$, then $f(\lambda^\dagger(\mathbf{H})) < 0$. Since the function $f(\lambda)$ is decreasing, this implies $f(\lambda^*) < 0$, and hence, equality does not hold in (7.23).

Let us prove the second part of the proposition. If $\lambda^\dagger(\mathbf{H}) = 0$, then the entropy is continuous on the set $\mathcal{K}_{\mathbf{H}}(h)$ by the observation in Wehrl [174]. It follows also from the implication (1) $\Rightarrow$ (2) in the below Proposition 7.3.10. To prove the converse implication, consider the sequence of states

$$\left(\sigma_n = (1 - q_n)|1\rangle_{\mathbb{H}}\langle 1| + \frac{q_n}{n}\sum_{k=2}^{n+1} |k\rangle_{\mathbb{H}}\langle k|\right)_{n=1}^{+\infty}$$

where

$$\left(q_n = (h - h_m(\mathbf{H}))\left(n^{-1}\sum_{k=2}^{n+1} h_k - h_m\right)^{-1}\right)_{n=1}^{+\infty}$$

is a sequence of positive numbers that converges to 0, where we assume that n is sufficiently large so that $q_n < 1$. Since the sequence $(\sigma_n)_{n=1}^{+\infty}$ lies in $\mathcal{K}_{\mathbf{H}}(h)$ and converges to the pure state $|1\rangle_{\mathbb{H}}\langle 1|$, continuity of $H(\rho)$ on the set $\mathcal{K}_{\mathbf{H}}(h)$ implies the convergence of the sequence $(H(\sigma_n))_{n=1}^{+\infty}$, to 0, where

$$H(\sigma_n) = h_2(q_n) + q_n \log n = h_2(q_n) + \frac{(h - h_m)\log n}{n^{-1}\sum_{k=2}^{n+1} h_k - h_m}.$$

By the obvious estimation $n^{-1}1\sum_{k=2}^{n+1} h_k \le h_{n+1}$, it follows that the sequence $(\nu_n = h_{n+1}\log n)_{n\in\mathbb{N}}$ converges to zero. Therefore, for arbitrary $\lambda > 0$, we have

$$\mathrm{tr}\big[\exp(-\lambda\mathbf{H})\big] = \sum_{n=0}^{+\infty} \exp(-\lambda h_{n+1}) = \sum_{n=1}^{+\infty} n^{-\frac{\lambda}{\nu_n}} < +\infty$$

and hence, $\lambda^{\dagger}(\mathbf{H}) = 0$. This proves the proposition. □

Let

$$h_*(\mathbf{H}) = \begin{cases} \frac{\mathrm{tr}[\mathbf{H}\exp(-\lambda^{\dagger}(\mathbf{H})\mathbf{H})]}{\mathrm{tr}[\exp(-\lambda^{\dagger}(\mathbf{H})\mathbf{H})]}, & \text{if } \mathrm{tr}[\exp(-\lambda^{\dagger}(\mathbf{H})\mathbf{H})] < +\infty; \\ +\infty, & \text{otherwise.} \end{cases} \tag{7.24}$$

Let $F_{\mathbf{H}}(\cdot) : [h_m(\mathbf{H}), +\infty[\to [0, +\infty]$ be the function defined by

$$F_{\mathbf{H}}(h) = \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho), \quad h \in [h_m(\mathbf{H}), +\infty[. \tag{7.25}$$

The following proposition (also due originally to Shirokov [141]) give precise calculation of $F_{\mathbf{H}}(h)$ for the range $h \le h_*(\mathbf{H})$ as well as for $h > h_*(\mathbf{H})$.

Proposition 7.3.7. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on the Hilbert space* $\mathbb{H}$ *and* h *be a positive number such that* $h > h_m(\mathbf{H})$*. Then:*
1. *If* $h \le h_*(\mathbf{H})$*, then*

$$F_{\mathbf{H}}(h) := \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) = \lambda_* h + \log\big(\mathrm{tr}\big[\exp(-\lambda_*\mathbf{H})\big]\big),$$

where $\lambda_* = \lambda_*(\mathbf{H}, h) \ge \lambda^{\dagger}(\mathbf{H})$ *is uniquely defined by the equation*

$$\mathrm{tr}\big[\mathbf{H}\exp(-\lambda\mathbf{H})\big] = h\,\mathrm{tr}\big[\exp(-\lambda\mathbf{H})\big], \tag{7.26}$$

and there exists the Gibbs state $\Gamma(\mathcal{K}_{\mathbf{H}}(h)) = \frac{\mathrm{tr}[\exp(-\lambda^*\mathbf{H})]}{\exp(-\lambda^*\mathbf{H})}$ *of the set* $\mathcal{K}_{\mathbf{H}}(h)$*.*

2. *If $h > h_*(\mathbf{H})$, then*

$$F_{\mathbf{H}}(h) := \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) = \lambda^\dagger(\mathbf{H})h + \log(\mathrm{tr}[\exp(-\lambda^\dagger(\mathbf{H})\mathbf{H})])$$

and there exists no $\rho \in \mathcal{K}_{\mathbf{H}}(h)$ such that $H(\rho) = F_{\mathbf{H}}(h)$. In all cases,

$$F_{\mathbf{H}}(h) := \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) = \inf_{\lambda^\dagger(\mathbf{H})<\lambda<+\infty} (\lambda h + \log(\mathrm{tr}[\exp(-\lambda\mathbf{H})])).$$

Proof. From the proof in Proposition 7.3.7, we observe that

$$F_{\mathbf{H}}(h) := \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) = \inf_{\lambda^\dagger(\mathbf{H})<\lambda<+\infty} (\lambda h + \log(\mathrm{tr}[\exp(-\lambda\mathbf{H})])).$$

The general expression for $F_{\mathbf{H}}(h) := \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho)$ can be deduced from the previous observation by noting that the infimum in this expression is achieved at λ_* if $h \leq h_*(\mathbf{H})$ and at $\lambda^\dagger(\mathbf{H})$ if $h \geq h_*(\mathbf{H})$. This proves the proposition. □

Example 7.3. As a classical counterpart for part (2) of Proposition 7.3.7, we demonstrate that Shannon entropy is continuous on a closed convex set $\mathcal{A}$ of classical probability distributions if and only if $\mathcal{A} \subset \mathcal{K}_{\{hi\},h}$ for a particular nondecreasing sequence $(h_i)_{i=1}^{+\infty}$ of nonnegative numbers that tends to $+\infty$ such that $\lambda^\dagger((h_i)_{i=1}^{+\infty}) = 0$ and $h > 0$, where

$$\lambda^\dagger((x_i)_{i=1}^{+\infty}) = \inf\left\{\lambda \mid \sum_i \exp(-\lambda h_i) < +\infty\right\}.$$

To prove this assertion, it is sufficient to prove the "only if" part, since the converse assertion follows from the similar assertion for the von Neumann entropy (see the previous section). Since the function $(x_i)_{i=1}^{+\infty} \mapsto S((x_i)_{i=1}^{+\infty})$ is finite on the closed convex set $\mathcal{A}$, it is bounded on this set and the classical analog of Corollary 5 in Shirokov [141] shows that the set $\mathcal{A}$ is compact. By the Dini's lemma, the condition $\|(x_i)_{i=1}^{+\infty}\|_1 = 1$ and the continuity of the function $(x_i)_{i=1}^{+\infty} \mapsto S((x_i)_{i=1}^{+\infty})$ imply uniform convergence of the series $\sum_i x_i$ and $\sum_i x_i(-\log x_i)$ on the set $\mathcal{A}$. Hence, there exists a sequence $(y_i)_{i=1}^{+\infty}$ of positive numbers tending to $+\infty$ such that

$$\sup_{(x_i)\in\mathcal{A}} \sum_i y_i x_i < +\infty \quad \text{and} \quad \sup_{(x_i)\in\mathcal{A}} \sum_i y_i x_i(-\log x_i) < +\infty. \tag{7.27}$$

Let $\mathcal{B}$ be the image of the set $\mathcal{A}$ under the map $(x_i)_{i=1}^{+\infty} \mapsto (y_i x_i)_{i=1}^{+\infty}$. It follows from (7.27) that $\mathcal{B}$ is a convex bounded subset of the positive cone of the space l_1 and that the extended Shannon entropy defined by the formula

$$S((x_i)_{i=1}^{+\infty}) = \|(x_i)_{i=1}^{+\infty}\|_{l^1} S\left(\frac{(x_i)_{i=1}^{+\infty}}{\|(x_i)_{i=1}^{+\infty}\|_{l^1}}\right)$$

is bounded on the set $\mathcal{B}$, where $\|(x_i)_{i=1}^{+\infty}\|_{l^1} = \sum_{i=1}^{+\infty}|x_i|$. By the classical analog of Lemma 2 in Shirokov [145], there exists a sequence $(h_i)_{i=1}^{+\infty}$ of nonnegative numbers $(h_i)_{i=1}^{+\infty}$ such that $\lambda^\dagger((h_i)_{i=1}^{+\infty}) < +\infty$ and $\sup_{(x_i)\in\mathcal{A}} h_i y_i x_i < +\infty$. It is easy to see that $\lambda^\dagger((h_i y_i)_{i=1}^{+\infty}) = 0$.

The following result (due originally to Shirokov [141]) investigates the behaviors of the function $F_{\mathbf{H}}(\cdot)$ defined in (7.25).

Proposition 7.3.8. *The function $F_{\mathbf{H}}(\cdot)$ has the following properties:*

1. *The function $F_{\mathbf{H}} : [h_m(\mathbf{H}), +\infty[\to [0, +\infty]$ is continuous and increasing such that $F_{\mathbf{H}}(h_m(\mathbf{H})) = \log(\dim(\mathbb{H}_m(\mathbf{H})))$ and $\lim_{h\to+\infty} F_{\mathbf{H}}(h) = +\infty$.*
2. *The function $F_{\mathbf{H}}(h)$ has a continuous derivative*

$$\frac{dF_{\mathbf{H}}(h)}{dh} = \begin{cases} \lambda^*(\mathbf{H}, h), & h \in]h_m(\mathbf{H}), h^*(\mathbf{H})[\\ \lambda^\dagger(\mathbf{H}), & h \in [h_*(\mathbf{H}), +\infty[\end{cases}$$

such that

$$\left.\frac{dF_{\mathbf{H}}(h)}{dh}\right|_{h=h_m(\mathbf{H})} = \lim_{h\downarrow h_m(\mathbf{H})} \frac{dF_{\mathbf{H}}(h)}{dh} = +\infty \quad \textit{and} \quad \lim_{h\to+\infty} \frac{dF_{\mathbf{H}}(h)}{dh} = \lambda^\dagger(\mathbf{H}).$$

3. *$F_{\mathbf{H}}(\cdot)$ is strictly concave on the interval $[h_m(\mathbf{H}), h_*(\mathbf{H})[$ and linear on $[h_*(\mathbf{H}), +\infty[$ if $h_*(\mathbf{H}) < +\infty$.*

Proof. As in the proof of Proposition 7.3.7, we will assume that the $\mathfrak{H}$-operator takes the form $\mathbf{H} = \sum_{k=1}^{+\infty} h_k |k\rangle_{\mathbb{H}}\langle k|$, where $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$ is an orthonormal basis in the space $\mathbb{H}$ and $(h_k)_{k=1}^{+\infty}$ is a nondecreasing sequence of positive numbers converging to the infinity. Let $d = \dim(\mathbb{H}_m(\mathbf{H}))$ so that $h_k = h_m(\mathbf{H})$ for $k = 1, 2, \ldots d$ and $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{d}$ is a basis of $\mathbb{H}_m(\mathbf{H})$.

The general expression for $F_{\mathbf{H}}(h) = \sup_{\rho\in\mathcal{K}_{\mathbf{H},h}} H(\rho)$ can be deduced from the previous observation by noting that the supremum in this expression is achieved at $\lambda^*(\mathbf{H}, h)$ if $h \le h_*(\mathbf{H})$ and at $\lambda^\dagger(\mathbf{H})$ if $h \ge h_*(\mathbf{H})$.

The proof of the properties of the function $F_{\mathbf{H}}(h)$ is based on the implicit function theorem (see, e. g., Rudin [133] and Halmos [57]) and is presented below. We first note that by lower semicontinuity of the von Neumann entropy

$$\lim_{h\to+\infty} F_{\mathbf{H}}(h) = \lim_{h\to+\infty} \sup_{\rho\in\mathcal{K}_{\mathbb{H}}(h)} H(\rho) = \sup_{\rho\in\mathcal{S}(\mathbb{H})} H(\rho) = +\infty$$

for arbitrary value of $\lambda^\dagger(\mathbf{H})$, since $\overline{\bigcup_{h\in\mathbb{R}} \mathcal{K}_{\mathbf{H}}(h)} = \mathcal{S}(\mathbb{H})$. Consider the function

$$g(\lambda, h) = \sum_{k=1}^{+\infty} (h_k - h)\exp(-\lambda h_k).$$

By using the theorem about series depending on parameters, it is easy to see that this function is differentiable at any point (λ, h) with $\lambda > \lambda^\dagger(\mathbf{H})$ and

$$\frac{\partial g(\lambda,h)}{\partial \lambda} = \sum_{k=1}^{+\infty} h_k(h-h_k)\exp(-\lambda h_k), \qquad \frac{\partial g(\lambda,h)}{\partial h} = -\sum_{k=1}^{+\infty} \exp(-\lambda h_k). \tag{7.28}$$

By the observation in the proof of Proposition 7.3.7 for each $h \in\,]h_m(\mathbf{H}), h_*(\mathbf{H})[$, there exists the unique $\lambda^* = \lambda^*(h) > \lambda^\dagger(\mathbf{H})$ such that $g(\lambda^*(h), h) = 0$. It follows from (7.28):

$$\left.\frac{\partial g(\lambda,h)}{\partial \lambda}\right|_{\lambda=\lambda^*(h)} = \sum_{k=1}^{+\infty} h_k(h-h_k)\exp(-\lambda^*(h)h_k) < 0.$$

By the implicit function theorem, the function $\lambda^*(h)$ is differentiable on the open interval $]h_m(\mathbf{H}), h_*(\mathbf{H})[$ and

$$\begin{aligned}\frac{d\lambda^*(h)}{dh} &= \left(\frac{\partial g(\lambda,h)}{\partial \lambda}\right)^{-1}\frac{\partial g(\lambda,h)}{\partial h}\\ &= \left(\sum_{k=1}^{+\infty} h_k(h-h_k)\exp(-\lambda^*(h)h_k)\right)^{-1}\sum_{k=1}^{+\infty}\exp(-\lambda^*(h)h_k) < 0.\end{aligned} \tag{7.29}$$

Expression (7.22) implies that

$$F_{\mathbf{H}}(h) = \lambda^*(h)h + \log\left(\sum_{k=1}^{+\infty}\exp(-\lambda^*(h)h_k)\right), \quad \forall h \in\,]h_m(\mathbf{H}), h_*(\mathbf{H})]. \tag{7.30}$$

By direct derivative calculation, we obtain

$$\frac{dF_{\mathbf{H}}(h)}{dh} = \frac{d}{dh}\left(\lambda^*(h)h + \log\left(\sum_{k=1}^{+\infty}\exp(-\lambda^*(h)h_k)\right)\right) = \lambda^*(h), \tag{7.31}$$

where the equality $g(\lambda^*(h), h) = 0$ was used. This and equation (7.29) imply

$$\frac{d^2F_{\mathbf{H}}(h)}{dh^2} = \frac{d\lambda^*(h)}{dh} < 0,$$

which shows strict concavity of the function $F_{\mathbf{H}}(h)$ on the open interval $]h_m(\mathbf{H}), h_*(\mathbf{H})[$. Suppose $h_*(\mathbf{H}) < +\infty$. If $h > h_*(\mathbf{H})$, then by the proved part of Proposition 7.3.7,

$$F_{\mathbf{H}}(h) = \lambda^\dagger(\mathbf{H})h + \log\left(\sum_{k=1}^{+\infty}\exp(-\lambda^\dagger(\mathbf{H})h_k)\right) \tag{7.32}$$

is a linear function and

$$\frac{dF_{\mathbf{H}}(h)}{dh} = \lambda^\dagger(\mathbf{H}). \tag{7.33}$$

If $h = h_*(\mathbf{H})$, then by the observation in the proof of Proposition 7.3.7, $\lambda^*(h) = \lambda^\dagger(\mathbf{H})$, and hence, representations (7.30) and (7.32) coincide in this case. To show smoothness of the function $F_{\mathbf{H}}(\cdot)$ at the point $h_*(\mathbf{H})$, note that $\lambda^*(h) \to \lambda^\dagger(\mathbf{H})$ as $h \uparrow h_*(\mathbf{H})$. Indeed, by (7.29) the function $\lambda^*(h)$ is decreasing on the open interval $[h_m(\mathbf{H}), h_*(\mathbf{H})[$ and for arbitrary $\lambda > \lambda^\dagger(\mathbf{H})$ there exists

$$h_\lambda = \left(\sum_{k=1}^{+\infty} \exp(-\lambda h_k)\right)^{-1} \sum_{k=1}^{+\infty} h_k \exp(-\lambda h_k)$$

such that $\lambda = \lambda^*(h_\lambda)$.

Thus, equations (7.30), (7.31), (7.32) and (7.33) together imply

$$\lim_{h \uparrow h_*(\mathbf{H})} F_{\mathbf{H}}(h) = F_{\mathbf{H}}(h_*(\mathbf{H})) \quad \text{and} \quad \lim_{h \uparrow h_*(\mathbf{H})} \frac{dF_{\mathbf{H}}(h)}{dh} = \left.\frac{dF_{\mathbf{H}}(h)}{dh}\right|_{h=h_*(\mathbf{H})+},$$

and hence, the function $F_{\mathbf{H}}(h)$ has a continuous derivative at the point $h_*(\mathbf{H})$. To prove right continuity of the function $F_{\mathbf{H}}(\cdot)$ at the point $h_m(\mathbf{H})$, note first that

$$\lim_{h \downarrow h_m(\mathbf{H})} \lambda^*(h) = +\infty. \tag{7.34}$$

Indeed, by (7.29) the function $\lambda^*(h)$ is decreasing on the open interval

$]h_m(\mathbf{H}), h_*(\mathbb{H})[$, and hence, there exists $\lambda_m = \lim_{h \downarrow h_m(\mathbf{H})} \lambda^*(h)$. If $\lambda_m < +\infty$ then by passing to the limit as $h \downarrow h_m(\mathbf{H})$ in the identity

$$\sum_{k=1}^{+\infty} h_k \exp(-\lambda^*(h) h_k) = h \sum_{k=1}^{+\infty} \exp(-\lambda^*(h) h_k),$$

valid for all h in the open interval $]h_m(\mathbf{H}), h_*(\mathbf{H})[$, we obtain a contradiction.

Let $d = \dim(\mathbb{H}_m(\mathbf{H}))$. It is easy to see that

$$P(h) = \log\left(\sum_{k=1}^{+\infty} \exp(-\lambda^*(h) h_k)\right) = -\lambda^*(h) h_m(\mathbf{H}) + Q(h), \tag{7.35}$$

where $Q(h) = \log(d + \sum_{k>d}^{+\infty} \exp(-\lambda^*(h)(h_k - h_m(\mathbf{H}))))$ is a nondecreasing function on the open interval $]h_m(\mathbf{H}), h_*(\mathbf{H})[$ tending to $\log(d)$ as $h \downarrow h_m(\mathbf{H})$. Since the function $F_{\mathbf{H}}(\cdot)$ is obviously nonnegative and nondecreasing on the interval $[h_m(\mathbf{H}), +\infty[$, there exists $\lim_{h \downarrow h_m(\mathbf{H})} F_{\mathbf{H}}(h) \geq F_{\mathbf{H}}(h_m(\mathbf{H}))$. This, Lemma 7.3.1, and (7.35) together imply that there exists

$$\lim_{h \downarrow h_m(\mathbf{H})} \lambda^*(h)(h - h_m(\mathbf{H})) = C < +\infty$$

and that

$$\lim_{h \downarrow h_m(\mathbf{H})} F_{\mathbf{H}}(h) = C + \log(d) = C + F_{\mathbf{H}}(h_m(\mathbf{H})).$$

Thus, to prove right continuity of the function $F_{\mathbf{H}}(\cdot)$ at the point $h_m(\mathbf{H})$ it is sufficient to show that $C = 0$. This can be done by proving that

$$\int_{h_m(\mathbf{H})}^{h''} \lambda(h)dh = \lim_{h' \downarrow h_m(\mathbf{H})} \int_{h'}^{h''} \lambda^*(h)dh < +\infty, \tag{7.36}$$

for some $h'' > h_m(\mathbf{H})$. Indeed, finiteness of this integral and the assumption $C > 0$ imply finiteness of the integral $\int_{h_m(\mathbf{H})}^{h''} (h - h_m(\mathbf{H}))^{-1} dh$.

It is easy to see that

$$\frac{dP(h)}{dh} = -h\frac{d\lambda^*(h)}{dh} \quad \text{and hence} \; -\frac{d\lambda^*(h)}{dh}(h - h_m(\mathbf{H})) = \frac{dQ(h)}{dh}.$$

By direct integration, we obtain

$$Q(h'') - Q(h') = \lambda^*(h')(h' - h_m(\mathbf{H})) - \lambda(h'')(h'' - h_m(\mathbf{H})) + \int_{h'}^{h''} \lambda^*(h)dh.$$

This and the mentioned before existence of $\lim_{h \downarrow h_m(\mathbf{H})} Q(h) = \log d$ and of $\lim_{h \downarrow h_m(\mathbf{H})} \lambda^*(h')(h' - h_m(\mathbf{H})) = C < +\infty$ imply (7.36). By the above observation,

$$\frac{F_{\mathbf{H}}(h) - F_{\mathbf{H}}(h_m(\mathbf{H}))}{h - h_m(\mathbf{H})} \geq \lambda^*(h), \quad \forall h > h_m(\mathbf{H}),$$

and hence, (7.34) implies $\frac{dF_{\mathbf{H}}(h)}{dh}|_{h=h_m(\mathbf{H})+} = +\infty$.

This proves the proposition. □

Proposition 7.3.9. *Let σ be a state with finite von Neumann entropy $H(\sigma)$. If $\lambda_\dagger(\sigma) < 1$, then $\lim_{n\to+\infty} H(\rho_n) = H(\sigma)$ for arbitrary sequence $(\rho_n)_{n=1}^{+\infty}$ of states H-converging to the state σ. If $\lambda_\dagger(\sigma) = 1$, then for arbitrary $h \geq H(\sigma)$ there exists a sequence $(\rho_n)_{n=1}^{+\infty}$ of states with finite support H-converging to the state σ such that $\lim_{n\to+\infty} H(\rho_n) = h$.*

Proof. Let $\lambda_\dagger(\sigma) < 1$. Then Lemma 7.3.5 implies

$$\frac{H(\rho_n \| (\mathrm{tr}[\sigma^\lambda])^{-1}\sigma^\lambda) - \lambda H(\rho_n\|\sigma)}{1-\lambda} = \frac{\log(\mathrm{tr}[\sigma^\lambda])}{1-\lambda} - H(\rho_n) \tag{7.37}$$

for all $\lambda > \lambda_\dagger(\sigma)$. Suppose $\liminf_{n\to+\infty} H(\rho_n) - H(\sigma) = \Delta > 0$. Since the first term in the right-hand side of (7.37) tends to $H(\sigma)$ as $\lambda \to 1$, there exists $\lambda' < 1$ such that the right-hand side of (7.37) is less than $-\Delta/2$ for this λ' and sufficiently large n. While by nonnegativity of the relative entropy, the left-hand side of (7.37) is greater than $-\frac{\lambda' H(\rho_n\|\sigma)}{1-\lambda'}$, which tends to zero as $n \to +\infty$. Let $\lambda_\dagger(\sigma) = 1$ and let $h > H(\sigma)$. Without loss of generality, we may assume that σ is a full rank state so that $\mathbf{H} = -\log\sigma$ is an

$\mathfrak{H}$-operator such that $\lambda^{\dagger}(\mathbf{H}) = \lambda_{\dagger}(\sigma) = 1$ and $h_*(\mathbf{H}) = H(\sigma) < +\infty$. By Proposition 7.3.7, $\sup_{\rho\in\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) = h$ for all $h > h_*(\mathbf{H})$. For given $h > h_*(\mathbf{H})$ in the proof of Proposition 7.3.7, the sequence $(\rho_n)_{n=1}^{+\infty}$ of states converging to the state $\rho_*(\mathcal{K}_{\mathbf{H}}(h)) = \sigma$ was constructed. By this construction,

$$\lim_{n\to\infty} H(\rho_n) = \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(h)} H(\rho) = h \quad \text{and} \quad \lim_{n\to+\infty} H(\rho_n\|\sigma) = 0.$$

This proves the proposition. □

We conclude this subsection by summarizing some results regarding the von Neumann entropy $H(\cdot)$ on the compact subset $\mathcal{K}_{\mathbf{H}}(h)$ of $\mathcal{S}(\mathbb{H})$, where $\mathbf{H}$ is an $\mathfrak{H}$-operator on $\mathbb{H}$.

- $H(\cdot)$ is finite on the set $\mathcal{K}_{\mathbf{H}}(h)$ if and only if $\lambda^{\dagger}(\mathbf{H}) < +\infty$;
- $H(\cdot)$ is continuous on the set $\mathcal{K}_{\mathbf{H}}(h)$ if and only if $\lambda^{\dagger}(\mathbf{H}) = 0$;
- $H(\cdot)$ is finite on a closed convex set $\mathcal{A}$ if and only if there exists an $\mathfrak{H}$-operator $\mathbf{H}$ with $\lambda^{\dagger}(\mathbf{H}) < +\infty$ such that $\mathcal{A} \subseteq \mathcal{K}_{\mathbf{H}}(h)$.

7.3.3 $H(\cdot)$ on an arbitrary closed subset $\mathcal{K}$

The following result is devoted to the question of continuity of the entropy on arbitrary subsets of states.

Proposition 7.3.10. *Let $\mathcal{K}$ be an arbitrary closed subset of $\mathcal{S}(\mathbb{H})$. The following properties are equivalent:*

1. *$\mathcal{K} \subseteq \mathcal{K}_{\mathbf{H}}(h)$ for some positive number h and $\mathfrak{H}$-operator $\mathbf{H}$ with $\lambda^{\dagger}(\mathbf{H}) = 0$;*
2. *The von Neumann entropy $H(\cdot)$ is continuous on the set $\mathcal{K}$ and there exists a state σ in $\mathcal{S}(\mathbb{H})$ such that the relative entropy $H(\cdot\|\sigma)$ is continuous and bounded on the set $\mathcal{K}$;*
3. *There exists a $\mathfrak{H}$-operator $\mathbf{H}$ with $\lambda^{\dagger}(\mathbf{H}) < +\infty$ such that the linear function $\mathrm{tr}[\cdot\mathbf{H}]$ is continuous and bounded on the set $\mathcal{K}$.*

If equivalent properties (1)–(3) hold, then the $\mathfrak{H}$-operators $\mathbf{H}$ and the state σ can be chosen in such a way that $\mathrm{tr}[\sigma\mathbf{H}] < +\infty$, $\mathbf{H} = -\log\sigma$ and $H(\sigma) < +\infty$.

Proof. (2) ⇒ (3). Since every continuous function is finite, we have from Definition 8.1.3 that

$$H(\rho\|\sigma) = -H(\rho) + \mathrm{tr}[\rho(-\log\sigma)], \quad \rho \in \mathcal{K}, \tag{7.38}$$

since $\mathrm{tr}[\rho] = \mathrm{tr}[\sigma] = 1$. By Proposition 3a (7.3.10), the set $\mathcal{K}$ is compact, and hence, the entropy is bounded on $\mathcal{K}$. Thus, the conditions of (2) and (7.38) imply continuity and boundedness of the function $\mathrm{tr}[\rho(-\log\sigma)](= H(\rho\|\sigma) + H(\rho))$ on the set $\mathcal{K}$, since both

$H(\cdot\|\sigma)$ and $H(\cdot)$ are bounded and continuous on $\mathcal{K}$. Hence, (3) holds with $\mathbf{H} = -\log\sigma$, which is an $\mathfrak{H}$-operator with $\lambda^{\dagger}(\mathbf{H}) < +\infty$ as shown in Example 7.2.

(3) $\Rightarrow$ (2). For given $\lambda > \lambda^{\dagger}(\tilde{\mathbf{H}})$, let $\sigma = (\mathrm{tr}[\exp(-\lambda\tilde{\mathbf{H}})])^{-1}\exp(-\lambda\tilde{\mathbf{H}})$ be a state in $\mathcal{S}(\mathbb{H})$ with finite entropy. Then (3) means continuity and boundedness of the function $\mathrm{tr}[\rho(-\log\sigma)]$ on the set $\mathcal{K}$. By lower semicontinuity of the entropy and of the relative entropy this and (7.38) imply continuity and boundedness of the functions $H(\cdot)$ and $H(\cdot\|\sigma)$ on the set $\mathcal{K}$.

(1) $\Rightarrow$ (3). By the assumption $\sum_k \exp(-\lambda h_k) < +\infty$ for all $\lambda > 0$, and hence, $\sum_k h_k \exp(-\lambda h_k) < +\infty$ for all $\lambda > 0$. This implies existence of a sequence $(\lambda_k)_{k=1}^{+\infty}$ of positive numbers monotonously converging to zero and such that $\sum_k h_k \exp(-\lambda_k h_k) < +\infty$. This sequence can be constructed as follows. For arbitrary natural m, let $N(m)$ be the minimal number such that

$$\sum_{k=N(m)}^{+\infty} \exp(-h_k/m) < 2^{-m}.$$

Consider a sequence $(\lambda_k)_{k=1}^{+\infty}$, where

$$\lambda_k = \left\{ \begin{array}{ll} 1, & k < N(2) \\ 1/m, & N(m) \le k < N(m+1), m \ge 2. \end{array} \right\}$$

It is easy to see that this sequence satisfies the above condition. Since $\mathrm{tr}[\rho\mathbf{H}] = \sum_k h_k \langle k|\rho|k\rangle_{\mathbb{H}} \le h$ for all $\rho \in \mathcal{K}$, the series $\sum_{k\in\mathbb{N}} \lambda_k h_k \langle k|\rho|k\rangle_{\mathbb{K}}$ converges uniformly on $\mathcal{K}$. This implies continuity of the function $\mathrm{tr}[\rho(-\log\sigma)]$, where

$$\sigma = \Big(\sum_k \exp(-\lambda_k h_k)\Big)^{-1} \sum_k \exp(-\lambda_k h_k)|k\rangle_{\mathbb{H}}\langle k|.$$

Note that the condition

$$\sum_k h_k \exp(-\lambda_k h_k) < +\infty \quad \text{implies } \mathrm{tr}[\sigma\mathbf{H}] < +\infty \text{ and } H(\sigma) < +\infty.$$

Thus, (3) holds with $\tilde{\mathbf{H}} = -\log\sigma$.

(3) $\Rightarrow$ (1). Let $\tilde{\mathbf{H}} = \sum_k \tilde{h}_k |k\rangle_{\mathbb{H}}\langle k|$, where $(|k\rangle_{\mathbb{H}})_{k\in\mathbb{N}}$ is an orthonormal basis in $\mathbb{H}$. Since (3) implies (2), Proposition 3a (7.3.10) implies compactness of $\mathcal{K}$. By the assumption, the series $\sum_k \tilde{h}_k \langle k|\rho|k\rangle_{\mathbb{H}}$ converges on the compact set $\mathcal{K}$ to the continuous function $\mathrm{tr}[\rho\tilde{\mathbf{H}}]$. By the Dini's lemma, it converges uniformly on $\mathcal{K}$. This implies existence of a sequence $(\lambda_k)_{k=1}^{+\infty}$ of positive numbers monotonously converges to $+\infty$ and such that $\sum_k \lambda_k \tilde{h}_k \langle k|\rho|k\rangle_{\mathbb{H}} \le h < +\infty$ for all $\rho \in \mathcal{K}$. It is easy to see that the $\mathfrak{H}$-operator $\mathbf{H} = \sum_k \lambda_k \tilde{h}_k |k\rangle_{\mathbb{H}}\langle k|$ has all the properties stated in (1).

The last assertion of the proposition follows from the above construction. This proves the proposition. □

We have the following observation from Propositions 7.3.7 and 7.3.10.

Corollary 7.3.11. *If* $\mathbf{H}$ *is an* $\mathfrak{H}$*-operator with* $\lambda^\dagger(\mathbf{H}) = 0$, *then there exists a state* σ *in* $\mathcal{S}(\mathbb{H})$ *and an* $\mathfrak{H}$*-operator* $\tilde{\mathbf{H}}$ *with* $\lambda^\dagger(\tilde{\mathbf{H}}) < +\infty$ *such that the relative entropy* $H(\cdot\|\sigma)$ *and linear functional* $\mathrm{tr}[\cdot\tilde{\mathbf{H}}]$ *are continuous on the set* $\mathcal{K}_{\mathbf{H}}(h)$.

Since the set $\mathcal{K}_{\mathbf{H},h}$ is convex by definition, Propositions 7.3.7 and 7.3.10 also provide the following result.

Corollary 7.3.12. *If the von Neumann entropy* $H(\cdot)$ *is continuous on the closed set* $\mathcal{K}$ *and there exists a state* $\sigma \in \mathcal{S}(\mathbb{H})$ *such that the relative entropy* $H(\cdot\|\sigma)$ *is continuous and bounded on the set* $\mathcal{K}$, *then the entropy is continuous on the set* $\overline{\mathrm{co}}(\mathcal{K})$.

Remark 7.1. The assumption of existence of the state σ in the statement (2) of Proposition 7.3.10 and in Corollary 7.3.12 is essential. Indeed, let $\mathcal{K}$ be the closed subset of all pure states in $\mathcal{S}(\mathbb{H})$. Then the von Neumann entropy $H(\cdot)$ is clearly continuous on this set $\mathcal{K}$, but it is not continuous on $\overline{\mathrm{co}}(\mathcal{K}) = \mathcal{S}(\mathbb{H})$. There exists compact countable set $\mathcal{K}$ of pure states such that $H(\cdot)$ is unbounded on the set $\overline{\mathrm{co}}(\mathcal{K})$.

The implication (3) $\Rightarrow$ (2) in Proposition 7.3.10 makes it possible to show continuity of the entropy on some nontrivial subsets of states, which will be used later.

Corollary 7.3.13. *Let* $\lambda \mapsto \mathfrak{U}_\lambda$ *be a continuous mapping from some compact set* Λ *into the set of all unitaries (antiunitaries) on* $\mathbb{H}$ *and let* ω *be a state in* $\mathcal{S}(\mathbb{H})$ *such that* $\mathfrak{U}_\lambda\omega\mathfrak{U}_\lambda^* = \omega$ *for all* $\lambda \in \Lambda$. *Then for arbitrary state* σ *such that* $\mathrm{tr}[\sigma(-\log\omega)] < +\infty$ *the functions* $H(\cdot)$ *and* $H(\cdot\|\omega)$ *are continuous on the set* $\overline{\mathrm{co}}(\{\mathfrak{U}\sigma\mathfrak{U}_\lambda^*\}\lambda \in \Lambda)$, *where* $\overline{\mathrm{co}}\{\cdots\}$ *denotes the smallest closed convex set containing* $\{\cdots\}$.

For an arbitrary orthonormal basis $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$ in $\mathbb{H}$, consider the expectation

$$\Pi_{\{|k\rangle_{\mathbb{H}}\}} : \rho \mapsto \sum_k \langle k|\rho|k\rangle_{\mathbb{H}}|k\rangle_{\mathbb{H}}\langle k|.$$

Note that the output states of $\Pi_{\{|k\rangle_{\mathbb{H}}\}}$ can be considered as classical states (probability distributions). So, we may call the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ classical projection of the set $\mathcal{K}$, corresponding to the basis $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$.

The following proposition shows, roughly speaking, that properties of sets of quantum states are closely related to the properties of classical projections of these sets.

Proposition 7.3.14. *Let* $\mathcal{K}$ *be an arbitrary closed subset of* $\mathcal{S}(\mathbb{H})$.
1. *The set* $\mathcal{K}$ *is compact if the set* $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ *is compact for at least one basis* $\{|k\rangle_{\mathbb{H}}\}$.
2. *If the set* $\mathcal{K}$ *is compact in* $\mathcal{S}(\mathbb{H})$, *then the set* $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ *is compact for arbitrary basis* $\{|k\rangle_{\mathbb{H}}\}$.
3. *The von Neumann entropy* $H(\cdot)$ *is bounded on the set* $\mathcal{K}$ *if it is bounded on the set* $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ *for at least one basis* $\{|k\rangle_{\mathbb{H}}\}$.

4. *If the von Neumann entropy $H(\cdot)$ is bounded on the set $\mathcal{K}$ and the set $\mathcal{K}$ is convex, then it is bounded on the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ for at least one basis $\{|k\rangle_{\mathbb{H}}\}$.*
5. *The von Neumann entropy $H(\cdot)$ is continuous on the set $\mathcal{K}$ if it is continuous on the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ for at least one basis $\{|k\rangle_{\mathbb{H}}\}$.*
6. *If the von Neumann entropy $H(\cdot)$ is continuous on the set $\mathcal{K}$ and there exists a state $\sigma \in \mathcal{S}(\mathbb{H})$ such that the relative entropy $H(\cdot\|\sigma)$ is continuous and bounded on the set $\mathcal{K}$, then the von Neumann entropy $H(\cdot)$ is continuous on the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ for at least one basis $\{|k\rangle_{\mathbb{H}}\}$.*

Proof. If the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ is compact, then by the compactness criterion for subsets of classical states for arbitrary $\epsilon > 0$ there exists $N_\epsilon > 0$ such that

$$\operatorname{tr}[\mathbf{P}_\epsilon \rho] = \sum_{k=1}^{N_\epsilon} \langle k|\rho|k\rangle_{\mathbb{H}} \geq 1 - \epsilon, \quad \forall \rho \in \mathcal{K},$$

where $\mathbf{P}_\epsilon = \sum_{k=1}^{N_\epsilon} |k\rangle_{\mathbb{H}}\langle k|$ is a finite rank projection. By the compactness criterion for subsets of $\mathcal{S}(\mathbb{H})$, this implies compactness of the set $\mathcal{K}$. If the set $\mathcal{K}$ is compact, then for arbitrary basis $\{|k\rangle_{\mathbb{A}}\}_{k\in\mathbb{N}}$ the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ is compact as an image of a compact set under a continuous mapping.

In the proof of the following statements, we will use the following identity:

$$H(\rho\|\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\rho)) = H(\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\rho)) - H(\rho), \tag{7.39}$$

valid for arbitrary state $\rho \in \mathcal{S}(\mathbb{H})$ with finite $H(\Pi_{\{\||k\rangle_{\mathbb{H}}\}}(\rho))$. If the von Neumann entropy $H(\cdot)$ is bounded on the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$, then it is bounded on the set $\mathcal{K}$ since identity (7.39) and nonnegativity of the relative entropy implies $H(\rho) \leq H(\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\rho))$ for arbitrary $\rho \in \mathcal{K}$. If the entropy is bounded on the convex set $\mathcal{K}$, then this set $\mathcal{K}$ is contained in the set $\mathcal{K}_{\mathbf{H},h}$ defined by a particular $\mathfrak{H}$-operator $\mathbf{H}$ with $\lambda^\dagger(\mathbf{H}) < +\infty$. Let $\{|k\rangle_{\mathbb{H}}\}_{k\in\mathbb{N}}$ be the basis of eigenvectors for the $\mathfrak{H}$-operator $\mathbf{H}$. Then $\Pi_{\{|k\rangle_{\{\mathbb{H}\}}\}}(\mathcal{K})$ also is contained in the set $\mathcal{K}_{\mathbf{H},h}$, and hence, the von Neumann entropy is bounded on the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ by Proposition 7.3.7. Suppose the von Neumann entropy $H(\cdot)$ is continuous on the set $\Pi\{|k\rangle_{\mathbb{H}}\}(\mathcal{K})$. Then the entropy is finite on this set, and by (7.39) it is finite on the set $\mathcal{K}$. Let ρ be a state in $\mathcal{K}$ and a sequence of states $(\rho_n)_{n\in\mathbb{N}}$ in $\mathcal{K}$ converging to ρ. By the assumption, lower semicontinuity of the relative entropy and (7.39), we have

$$\begin{aligned}
&\limsup_{n\to+\infty} H(\rho_n) \\
&\quad = \lim_{n\to+\infty} H(\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\rho_n)) - \liminf_{n\to+\infty} H(\rho_n\|\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\rho_n)) \\
&\quad \leq H(\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\rho)) - H(\rho\|\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\rho)) = H(\rho).
\end{aligned}$$

This and lower semicontinuity of the entropy imply $\lim_{n\to+\infty} H(\rho_n) = H(\rho)$.

If the entropy is continuous on the set $\mathcal{K}$ and there exists a state σ in $\mathcal{S}(\mathbb{H})$ such that the relative entropy $H(\rho\|\sigma)$ is continuous and bounded on the set $\mathcal{K}$ then by Proposition 7.3.10, the set $\mathcal{K}$ is contained in the set $\mathcal{K}_{\mathbf{H},h}$ defined by a particular $\mathfrak{H}$-operator $\mathbf{H}$ with $\lambda^{\dagger}(\mathbf{H}) = 0$. Let $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$ be the basis of eigenvectors for $\mathbf{H}$. Then $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ also is contained in the set $\mathcal{K}_{\mathbf{H},h}$, and hence, the von Neumann entropy $H(\cdot)$ is continuous on the set $\Pi_{\{|k\rangle_{\mathbb{H}}\}}(\mathcal{K})$ by Proposition 7.3.7. This proves the proposition. □

7.4 $H(\cdot)$ extended to $\mathfrak{T}_+(\mathbb{H})$

Recall that $\mathfrak{T}_+(\mathbb{H})$ is the positive cone of the Banach space of trace-class operators $\mathfrak{T}(\mathbb{H})$. That is,

$$\mathfrak{T}_+(\mathbb{H}) = \{\mathbf{A} \in \mathfrak{T}(\mathbb{H}) \mid \mathbf{A} \geq \mathbf{0}\}.$$

In this section, we extend von Neumann entropy $H(\cdot)$ and relative entropy $H(\cdot\|\cdot)$ from the space of quantum states $\mathcal{S}(\mathbb{H})$ to the positive cone of trace- class operators $\mathfrak{T}_+(\mathbb{H})$ on $\mathbb{H}$.

We first consider any real valued or extended real valued function F from its domain $\mathcal{S}(\mathbb{H})$ to $\mathfrak{T}_+(\mathbb{H})$, also denoted by F, via the following formula:

$$F(\mathbf{A}) = \operatorname{tr}[\mathbf{A}]F\left(\frac{\mathbf{A}}{\operatorname{tr}[\mathbf{A}]}\right), \quad \forall \mathbf{A} \in \mathfrak{T}_+(\mathbb{H}), \tag{7.40}$$

where $\frac{\mathbf{A}}{\operatorname{tr}[\mathbf{A}]} \in \mathcal{S}(\mathbb{H})$.

While we make no notational distinction before and after the extension of a function to $\mathfrak{T}_+(\mathbb{H})$, we note that F on the left-hand side of the above equation denotes a function with domain in $\mathfrak{T}_+(\mathbb{H})$, whereas F in the right-hand side denotes a function with domain in $\mathcal{S}(\mathbb{H})$.

Motivated by the definitions of von Neumann and relative entropies for quantum states, it is natural to extend these quantum entropies from $\mathcal{S}(\mathbb{H})$ to $\mathfrak{T}_+(\mathbb{H})$. By choosing $F(x) = \eta(x)$ for $x \geq 0$, we can define the extended von Neumann entropy $H(\cdot) : \mathfrak{T}_+(\mathbb{H}) \to [0, +\infty]$ as follows.

Definition 7.4.1. The von Neumann entropy of an operator $\mathbf{A} \in \mathfrak{T}_+(\mathbb{H})$ is defined as follows:

$$H(\mathbf{A}) = \operatorname{tr}[\eta(\mathbf{A})] - \eta(\operatorname{tr}[\mathbf{A}]). \tag{7.41}$$

Note that when $\mathbf{A} \in \mathcal{S}(\mathbb{H})$, then $\operatorname{tr}[\mathbf{A}] = 1$ and $\eta(\operatorname{tr}[\mathbf{A}]) = 0$. The extended von Neumann entropy reduces to the von Neumann entropy $H(\mathbf{A}) = \operatorname{tr}[\eta(\mathbf{A})]$.

Similar to the extension of $H(\cdot)$ to $\mathfrak{T}_+(\mathbb{H})$, classical entropy $H(\cdot) : \mathfrak{P}_1 \to [0, +\infty]$ can also be extended to the positive cone of the Banach space

$$(l_1)_+ := \left\{ (x_n)_{n=1}^{+\infty} \mid x_n > 0, \sum_{n=1}^{+\infty} x_n < +\infty \right\}$$

as follows. Let $(x_n)_{n=1}^{+\infty}$ be a sequence of real numbers in $(l_1)_+$. With a little abuse of notation of von Neumann entropy $H(\cdot)$, we define *classical entropy* $H(\cdot) : (l_1)_+ \to [0, +\infty]$ as

$$(x_n)_{n=1}^{+\infty} \mapsto H\big((x_n)_{n=1}^{+\infty}\big) := \sum_{n=1}^{+\infty} \eta(x_n) - \eta\left(\sum_{n=1}^{+\infty} x_n \right). \tag{7.42}$$

The restriction of the classical entropy to

$$\mathfrak{P}_1 = \left\{ (x_n)_{n=1}^{+\infty} \mid x_n > 0, \sum_{n=1}^{+\infty} x_n = 1 \right\} \subset (l_1)_+$$

reduces to the Shannon entropy defined earlier. This is because

$$H\big((x_n)_{n=1}^{+\infty}\big) = \sum_{n=1}^{+\infty} \eta(x_n) = \sum_{n=1}^{+\infty} -x_n \log x_n, \quad \forall (x_n)_{n=1}^{+\infty} \in \mathfrak{P}_1,$$

because $\eta(\sum_{n=1}^{+\infty} x_n) = \eta(1) = 0$.

With the exception of Shannon entropy $S(\cdot)$, the same notation $H(\cdot)$ has been used for von Neumann entropy, extended von Neumann entropy and classical entropy defined above. However, the type of entropies $H(\cdot)$ is referred to should be clear from its context or explicitly mentioned throughout the end of this book. The same also applies to $H(\cdot\|\cdot)$.

7.4.1 Properties of $H(\cdot)$ on $\mathfrak{T}_+(\mathbb{H})$

Using Definition 7.4.1 and well-known properties of the von Neumann entropy on $\mathcal{S}(\mathbb{H})$ (see Section 7.2), it is easy to obtain the following properties for the extended von Neumann entropy on $\mathfrak{T}_+(\mathbb{H})$. In particular, it is easy to show the nonnegativity, concavity and lower semicontinuity of the extended von Neumann entropy $H(\cdot)$ follows from those of the von Neumann entropy $H(\cdot)$ on $\mathcal{S}(\mathbb{H})$ (see Propositions 7.2.3, 7.2.6 and 7.2.7).

Proposition 7.4.2. *The following properties of the extended von Neumann entropy* $H(\cdot); \mathfrak{T}_+(\mathbb{H}) \to [0, +\infty]$ *hold:*

1. *For each* $\mathbf{A} \in \mathfrak{T}_+(\mathbb{H})$ *and* $\lambda \geq 0$,

$$H(\lambda\mathbf{A}) = \lambda H(\mathbf{A}), \quad \lambda \geq 0, \tag{7.43}$$

2. (Concavity) *For* $\lambda \in [0, 1]$ *and* $\mathbf{A}, \mathbf{B} \in \mathfrak{T}(\mathbb{H})$, *we have*

$$H\big(\lambda\mathbf{A} + (1-\lambda)\mathbf{B}\big) \geq \lambda H(\mathbf{A}) + (1-\lambda)H(\mathbf{B}). \tag{7.44}$$

3. (Lower semicontinuity) $H(\mathbf{A}) \le \liminf_{n\to+\infty} H(\mathbf{A}_n)$ *for any sequence* $(\mathbf{A}_n)_{n=1}^{+\infty}$ *in* $\mathfrak{T}(\mathbb{H})$ *such that* $\lim_{n\to+\infty} \mathbf{A}_n = \mathbf{A} \in \mathfrak{T}(\mathbb{H})$ *under* $\|\cdot\|_1$*-norm.*
4. (Strong subadditivity) *For composite Hilbert space* $\mathbb{H}_{123} = \mathbb{H}_1 \otimes \mathbb{H}_2 \otimes \mathbb{H}_3$ *and let* $\mathbf{A}_{123} \in \mathfrak{T}(\mathbb{H}_{123})$*. For* $i, j, k = 1, 2, 3$*, denote* $\mathrm{tr}_{\mathbb{H}_i \otimes \mathbb{H}_j}[\mathbf{A}_{123}]$ *by* $\mathbf{A}_k$ *and* $\mathrm{tr}_{\mathbb{H}_k}[\mathbf{A}_{123}]$ *by* $\mathbf{A}_{ij}$*. Then*

$$\begin{aligned} H(\mathbf{A}_{123}) + H(\mathbf{A}_2) &\le H(\mathbf{A}_{12}) + H(\mathbf{A}_{23}) \\ H(\mathbf{A}_1) + H(\mathbf{A}_2) &\le H(\mathbf{A}_{13}) + H(\mathbf{A}_{23}). \end{aligned}$$

5. (Monotonicity) *For* $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$*,*

$$\mathbf{A} \le \mathbf{B} \Rightarrow H(\mathbf{A}) \le H(\mathbf{B}), \tag{7.45}$$

and

$$H(\mathbf{A}) + H(\mathbf{B} - \mathbf{A}) \le H(\mathbf{B}) \le H(\mathbf{A}) + H(\mathbf{B} - \mathbf{A}) + \mathrm{tr}[\mathbf{B}] h_2\left(\frac{\mathrm{tr}[\mathbf{A}]}{\mathrm{tr}[\mathbf{B}]}\right), \tag{7.46}$$

where $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$*,* $\mathbf{A} \le \mathbf{B}$*, and* $h_2(x) = \eta(x) + \eta(1 - x)$*.*

Proof. 1. We first note that

$$\begin{aligned} H(\lambda\mathbf{A}) &= \mathrm{tr}[\eta(\lambda\mathbf{A})] - \eta(\mathrm{tr}[\lambda\mathbf{A}]) \\ &= -\,\mathrm{tr}[\lambda\mathbf{A}\log(\lambda\mathbf{A})] + \mathrm{tr}[\lambda\mathbf{A}]\log(\mathrm{tr}[\lambda\mathbf{A}]) \\ &= -\,\mathrm{tr}[\lambda\mathbf{A}(\log\lambda + \log(\mathbf{A}))] + \mathrm{tr}[\lambda\mathbf{A}](\log\lambda + \log(\mathrm{tr}[\mathbf{A}])) \\ &= -\lambda\,\mathrm{tr}[\mathbf{A}\log\mathbf{A}] + \lambda\,\mathrm{tr}[\mathbf{A}]\log(\mathrm{tr}[\mathbf{A}]) \\ &= \lambda\,\mathrm{tr}[\eta(\mathbf{A})] - \eta(\mathrm{tr}[\mathbf{A}]) = \lambda H(\mathbf{A}), \quad \forall \mathbf{A} \in \mathfrak{T}_+(\mathbb{H}) \text{ and } \lambda \ge 0. \end{aligned}$$

5. Consequently, by taking the concavity of the extended von Neumann entropy and (7.43), we have the following monotonicity property:

$$\mathbf{A} \le \mathbf{B} \Rightarrow H(\mathbf{A}) \le H(\mathbf{B}), \quad \forall \mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H}). \tag{7.47}$$

This completes the proof. □

We will also use the following lemmas.

Lemma 7.4.3. *Let* $(\mathbf{A}_n)_{n=1}^{+\infty}$ *and* $(\mathbf{B}_n)_{n=1}^{+\infty}$ *be sequences of operators in* $\mathfrak{T}_+(\mathbb{H})$ *converging in trace-norm* $\|\cdot\|_1$ *to operators* $\mathbf{A}_0$ *and* $\mathbf{B}_0$*, respectively. Then*

$$\lim_{n\to+\infty} H(\mathbf{A}_n + \mathbf{B}_n) = H(\mathbf{A}_0 + \mathbf{B}_0)$$

if and only if

$$\lim_{n\to+\infty} H(\mathbf{A}_n) = H(\mathbf{A}_0) \quad \text{and} \quad \lim_{n\to+\infty} H(\mathbf{B}_n) = H(\mathbf{B}_0).$$

Proof. ($\Rightarrow$) The implication follows immediately by first choosing $\mathbf{B}_n = \mathbf{0}$ for all n and then choosing $\mathbf{A}_n = \mathbf{0}$ for each $n \in \mathbb{N}$.

($\Leftarrow$) We now assume that $\lim_{n\to+\infty} H(\mathbf{A}_n) = H(\mathbf{A}_0)$ and $\lim_{n\to+\infty} H(\mathbf{B}_n) = H(\mathbf{B}_0)$. Then by lower semicontinuity of $H(\cdot)$,

$$H(\mathbf{A}_0) + H(\mathbf{B}_0) = \lim_{n\to+\infty} H(\mathbf{A}_n) + \lim_{n\to+\infty} H(\mathbf{B}_n) \le \lim_{n\to+\infty} H(\mathbf{A}_n + \mathbf{B}_n).$$

On the other hand, by convexity of $H(\cdot)$, we have

$$\lim_{n\to+\infty} H(\mathbf{A}_n + \mathbf{B}_n) \le \lim_{n\to+\infty} H(\mathbf{A}_n) + \lim_{n\to+\infty} H(\mathbf{B}_n) = H(\mathbf{A}_0) + H(\mathbf{B}_0).$$

This proves the lemma. □

Proposition 7.4.4. *Let $\{\mathbf{A}_i\}_{i=1}^n$ be a set of operators in $\mathfrak{T}_{\le 1}(\mathbb{H})$ and $\{\lambda_i\}_{i=1}^n$ be a set of positive real numbers such that $\sum_{i=1}^n \lambda_i = 1$, where $n \le +\infty$. Then*

$$\sum_{i=1}^n \lambda_i H(\mathbf{A}_i) \le H\left(\sum_{i=1}^n \lambda_i \mathbf{A}_i\right) \le \sum_{i=1}^n \lambda_i H(\mathbf{A}_i) + H(\{\lambda_i\}_{i=1}^n). \tag{7.48}$$

Proof. The first inequality in (7.48) follows from the concavity of the extended von Neumann entropy $H(\cdot)$. To show the second inequality, we first consider a special case where $\mathbf{A}_i = \mathbf{P}_i$ (a one-dimensional projection). In this case,

$$H\left(\sum_{i=1}^n \lambda_i \mathbf{A}_i\right) = \operatorname{tr}\left[\eta\left(\sum_{i=1}^n \lambda_i \mathbf{P}_i\right)\right] - \eta\left(\sum_{i=1}^n \lambda_i \operatorname{tr}[\mathbf{P}_i]\right) \le -\sum_{i=1}^n \lambda_i \log \lambda_i.$$

Inequality (7.48) then follows in the general case because if one decomposes ρ_i into one-dimensional spectral projections,

$$\rho_i = \sum_{k=1}^{+\infty} p_k^{(i)} \mathbf{Q}_k^{(i)},$$

then

$$\begin{aligned}
H\left(\sum_{i=1}^n \lambda_i \rho_i\right) &= H\left(\sum_{i,k} \lambda_i p_k^{(i)} \mathbf{Q}_k^{(i)}\right) \\
&\le -\sum_{i,k} \lambda_i p_k^{(i)} (\log \lambda_i + \log p_k^{(i)}) \\
&= -\sum_i \lambda_i \log \lambda_i - \sum_i \lambda - i \sum_k p_k^{(i)} \log p_k^{(i)} \\
&= -\sum_i \lambda_i \log \lambda_i + \sum_i \lambda_i H(\rho_i).
\end{aligned}$$

This proves the proposition. □

The following two corollaries follow immediately from Proposition 7.4.4.

Corollary 7.4.5. *Let $\{\mathbf{A}_i\}_{i=1}^n$ be a set of operators in $\mathfrak{T}_{\leq 1}(\mathbb{H})$ with finite $\sum_{i=1}^n \operatorname{tr}[\mathbf{A}_i]$. Then*

$$\sum_{i=1}^n H(\mathbf{A}_i) \leq H\left(\sum_{i=1}^n \mathbf{A}_i\right) \leq \sum_{i=1}^n \lambda_i H(\mathbf{A}_i) + H(\{\operatorname{tr}[\mathbf{A}_i]\}_{i=1}^n). \tag{7.49}$$

Corollary 7.4.6. *The extended von Neumann entropy $H(\mathbf{A})$ of an arbitrary operator $\mathbf{A} \in \mathfrak{T}_+(\mathbb{H})$ and the classical $H(\cdot)$ of the sequence of its diagonal values in any orthonormal basis $(|i\rangle_{\mathbb{H}})_{i=1}^{+\infty}$ of the space $\mathbb{H}$ are related as follows:*

$$H(\mathbf{A}) \leq H((\langle i|\mathbf{A}|i\rangle_{\mathbb{H}})_{i=1}^{+\infty}) \tag{7.50}$$

Corollary 7.4.7. *Let $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_{\leq 1}(\mathbb{H})$ and $\lambda \in]0,1[$. Then*

$$H(\lambda\mathbf{A} + (1-\lambda)\mathbf{B}) \leq \lambda H(\mathbf{A}) + (1-\lambda)H(\mathbf{B}) + h_2(\lambda), \tag{7.51}$$

where

$$n_2(\lambda) = \eta(\lambda) + \eta(1-\lambda) = -\lambda\log\lambda - (1-\lambda)\log(1-\lambda)$$

is the binary entropy.

The conclusion of the following proposition comes immediately from Lemma 7.4.6. The proof is therefore omitted.

Proposition 7.4.8. *Let $(|i\rangle_{\mathbb{H}})_{i=1}^{+\infty}$ be an orthonormal basis of a Hilbert space $\mathbb{H}$. Continuity of the von Neumann entropy $H(\cdot)$ on a set $\mathcal{A} \subset \mathfrak{T}_+(\mathbb{H})$ follows from the continuity of the classical entropy on the set*

$$\{\{\langle i|\mathbf{A}|i\rangle_{\mathbb{H}}\}_{i=1}^{+\infty} \mid \mathbf{A} \in \mathcal{A}\} \subset (l^1)_+,$$

where $(l^1)_+$ is the positive cone of the space l^1, i. e.,

$$(l^1)_+ = \left\{(c_n)_{n=1}^{+\infty} \mid c_n \geq 0, \sum_{n=1}^{+\infty} c_n < +\infty\right\}.$$

7.4.2 Discontinuity of extended quantum entropies

This subsection (written based on the results obtained in Shirokov [150]) deals with discontinuity of $H(\cdot) : \mathfrak{T}_+(\mathbb{H}) \to [0,+\infty]$ and $H(\cdot\|\cdot) : \mathfrak{T}_+(\mathbb{H}) \times \mathfrak{T}_+(\mathbb{H}) \to [-\infty,+\infty]$.

Since the extended von Neumann $H(\cdot) : \mathfrak{T}_+(\mathbb{H}) \to [0, +\infty]$ is lower semicontinuous,

$$H(\mathbf{A}_0) = \liminf_{n\to+\infty} H(\mathbf{A}_n) \le \limsup_{n\to+\infty} H(\mathbf{A}_n)$$

for any sequence $(\mathbf{A}_n)_{n=1}^{+\infty} \subset \mathfrak{T}_+(\mathbb{H})$ that converges to $\mathbf{A}_0 \in \mathfrak{T}(\mathbb{H})$ in the $\|\cdot\|_1$-norm. By the lower semicontinuity of the extended von Neumann entropy $H(\cdot)$, its discontinuity jumps for such a sequence can be characterized by the nonnegative value

$$dj(H(\mathbf{A}_n)) = \limsup_{n\to+\infty} H(\mathbf{A}_n) - H(\mathbf{A}_0),$$

where it is assumed that $djH(\mathbf{A}_n) = +\infty$ if $H(\mathbf{A}_0) = +\infty$. This value can be called the *entropy loss* corresponding to the sequence $(\mathbf{A}_n)_{n=1}^{+\infty}$.

We have the following simple but useful observation.

Proposition 7.4.9. *Let $(\mathbf{A}_n)_{n=1}^{+\infty} \subset \mathfrak{T}_+(\mathbb{H})$ be a sequence converging to an operator $\mathbf{A}_0$ in the $\|\cdot\|_1$-norm. Then*

$$dj(H(\mathbf{A}_n)) \le \limsup_{n\to+\infty} \operatorname{tr}[\mathbf{A}_n(-\log \mathbf{B}_n)] - \operatorname{tr}[\mathbf{A}_0(-\log \mathbf{B}_0)] \tag{7.52}$$

for any sequence $(\mathbf{B}_n)_{n=1}^{+\infty} \subset \mathfrak{T}_+(\mathbb{H})$ converging to an operator $\mathbf{B}_0$ in the $\|\cdot\|_1$-norm, where it is assumed that the right-hand side is equal to $+\infty$ if $\operatorname{tr}[\mathbf{A}_0(-\log \mathbf{B}_0)] = +\infty$.

Proof. We first note that if $\mathbf{B}_n = \mathbf{A}_n$ for all n, then (7.52) becomes trivial. It follows from (7.41) and (7.46) that

$$H(\mathbf{A}_n) + H(\mathbf{A}_n\|\mathbf{B}_n) + f(\mathbf{A}_n) - \operatorname{tr}[\mathbf{B}_n] = \operatorname{tr}[\mathbf{A}_n(-\log \mathbf{B}_n)] \tag{7.53}$$

for all $n \ge 0$, where $f(\mathbf{A}_n) = \eta(\operatorname{tr}[\mathbf{A}_n]) + \operatorname{tr}[\mathbf{A}_n]$. Hence,

$$\begin{aligned}&\limsup_{n\to+\infty} H(\mathbf{A}_n) + \liminf_{n\to+\infty} H(\mathbf{A}_n\|\mathbf{B}_n) + \lim_{n\to+\infty} (f(\mathbf{A}_n) - \operatorname{tr}[\mathbf{B}_n])\\ &\quad \le \limsup_{n\to+\infty} \operatorname{tr}[\mathbf{A}_n(-\log \mathbf{B}_n)].\end{aligned}$$

By subscribing equality (7.53) with $n = 0$ from this inequality and by using the lower semicontinuity of the relative entropy, we obtain (7.52). This proves the proposition. □

Let $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$ be an orthonormal basis in $\mathbb{H}$. For any state $\rho \in \mathcal{S}(\mathbb{H})$, we may consider the probability distribution $\pi(\rho) = \{\langle k|\rho|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$. It is well known from Corollary 7.4.6 that $H(\rho) \le S(\pi(\rho))$, where $S(\cdot)$ is the Shannon entropy, which is a lower semicontinuous function on the set of all countable probability distributions equipped with the l_1 metric. Proposition 7.4.9 shows that a similar relation hold for jumps of the entropy corresponding to converging sequences $(\rho_n)_{n=1}^{+\infty}$ and $(\pi(\rho_n))_{n=1}^{+\infty}$.

Corollary 7.4.10. *For any sequence $(\rho_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H})$ converging to a state ρ_0, we have*

$$djH(\rho_n) \leq dj(S(\pi(\rho_n))) = \limsup_{n\to+\infty} S(\pi(\rho_n)) - S(\pi(\rho_0)), \tag{7.54}$$

where it is assumed that $dj(S(\pi(\rho_n)))$ is equal to $+\infty$ if $S(\pi(\rho_0)) = +\infty$. Note that "=" holds in (7.54) if the sequence $(\rho_n)_{n=1}^{+\infty}$ consists of states diagonalizable in the basis $\{|k\rangle_{\mathbb{H}}\}_{k=1}^{+\infty}$.

8 Relative and conditional entropies

8.1 Quantum relative entropy

As a motivation for developing quantum relative entropy, we first recall absolute continuity of measures (see Halmos [57] and Rudin [133]) and Kullback–Leibier divergence (see Kullback–Leibier [101]) below.

Definition 8.1.1 (Absolute continuity and Radon–Nikodym derivative). Let P and Q be probability measures on a (Borel) measurable space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. The measure P is said to be *absolutely continuous* with respect to measure Q if, for all $A \in \mathcal{B}(\mathbb{X})$, $Q(A) = 0$, implies $P(A) = 0$. In this case, there exists a measurable function called *the Radon–Nikodym derivative* and denoted symbolically by $\frac{dP}{dQ}(x), x \in \mathbb{X}$, such that

$$P(A) = \int_A \frac{dP}{dQ}(x) Q(dx), \quad \forall A \in \mathcal{B}(\mathbb{X}). \tag{8.1}$$

Definition 8.1.2 (Kullback–Leibier divergence). Let P and Q be two probability measures defined on (Borel) measurable space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ such that P is absolutely continuous with respect to Q with the Radon–Nikodym derivative $\frac{dP}{dQ}(x)$. The Kullback–Leibier divergence of P with respect to Q is defined by

$$D_{KL}(P\|Q) = \int_{\mathbb{X}} \log\left(\frac{dP}{dQ}(x)\right) P(dx) \tag{8.2}$$

provided that the right-hand expression above exists.

Some special cases of Kullback–Leibier divergence are illustrated below. For discrete probability measures P and Q defined on the same probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the Kullback–Leibier divergence between P and Q is reduced to

$$D_{KL}(P\|Q) = \sum_{x \in \mathbb{R}} P(x) \log\left(\frac{P(x)}{Q(x)}\right). \tag{8.3}$$

For distribution P and Q for continuous random variables, the Kullback–Leibier divergence between P and Q becomes

$$D_{KL}(P\|Q) = \int_{-\infty}^{+\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx, \tag{8.4}$$

where $p(x)$ and $q(x)$ are probability densities of P and Q, respectively. Furthermore, if μ is any measure on $\mathbb{X}$ for which $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ exist (meaning that p and q are absolutely continuous with respect to μ), then the Kullback–Leibier divergence becomes

https://doi.org/10.1515/9783110788105-008

$$D_{KL}(P\|Q) = \int_{\mathbb{X}} p \log\left(\frac{p}{q}\right) d\mu. \tag{8.5}$$

In (classical) mathematical statistics, the Kullback–Leibier divergence (also called relative entropy) is a measure of how one probability distribution is different from a second, reference probability distribution, which can be used to characterize the relative (Shannon) entropy in classical information systems. In this case, log in both (8.4) and (8.5) is taken to be based 2 for the unit of bits.

In contrast to variation of information, Kullback–Leibier divergence is a distributionwise asymmetric measure, and thus does not qualify as a statistical metric of spread-it, and also does not satisfy the triangle inequality. In the simple case, a Kullback–Leibier divergence of 0 indicates that the two distributions in question are identical.

In the context of quantum information, the relative entropy $H(\rho\|\sigma)$ of states ρ and σ defined in (8.6) and (8.7) can be considered as a measure of divergence of these states, which generalize its classical analog of Kullback–Leibier divergence described above.

We first define relative entropy of quantum states for finite-dimensional Hilbert space $\mathbb{H}$ below.

Definition 8.1.3. Let $\mathbf{A}$ and $\mathbf{B}$ be two positive finite rank operators on the Hilbert space $\mathbb{H}$. We define the relative entropy $H(\mathbf{A}\|\mathbf{B})$ of $\mathbf{A}$ and $\mathbf{B}$ as

$$H(\mathbf{A}\|\mathbf{B}) = \begin{cases} \operatorname{tr}[\mathbf{A}\log\mathbf{A} - \mathbf{A}\log\mathbf{B} + \mathbf{B} - \mathbf{A}], & \text{if range}(\mathbf{A}) \subset \text{range}(\mathbf{B}) \\ +\infty, & \text{otherwise.} \end{cases} \tag{8.6}$$

We now define relative entropies of quantum states on infinite-dimensional Hilbert space $\mathbb{H}$ as follows.

Definition 8.1.4. Let ρ and σ be quantum states in $\mathcal{S}(\mathbb{H})$. The *relative entropy* $H(\rho\|\sigma)$ of states ρ and σ is defined by

$$H(\rho\|\sigma) = \begin{cases} \sum_i \langle i|\rho\log\rho - \rho\log\sigma\rangle_{\mathbb{H}}, & \text{if } \ker^{\perp}(\rho) \subset \ker^{\perp}(\sigma) \\ +\infty, & \text{otherwise,} \end{cases} \tag{8.7}$$

where $\{|i\rangle_{\mathbb{H}}\}_{i=1}^{+\infty} \subset \mathbb{H}$ is a basis of the eigenvectors of ρ.

Comparing the above two definitions, Definitions 8.1.3 and 8.1.4, we note that if $\mathbf{A}$ and $\mathbf{B}$ are quantum states on $\mathbb{H}$, then $\operatorname{tr}[\mathbf{B}] = \operatorname{tr}[\mathbf{A}] = 1$. Consequently, the two formulae (8.6) and (8.7) coincide.

We need the following Lieb concavity theorem to establish some properties of relative entropy $H(\cdot\|\cdot)$.

Lemma 8.1.5 (Lieb's concavity theorem [104]). *Let* $\mathbf{A}_1 \geq \mathbf{0}$ *and* $\mathbf{A}_2 \geq \mathbf{0}$ *and* $\mathbf{K}$ *is any operator. Then for* $0 \leq t \leq 1$, *and then*

$$(\mathbf{A}_1, \mathbf{A}_2) \mapsto \mathrm{tr}[\mathbf{K}\mathbf{A}_1^t\mathbf{K}^*\mathbf{A}_2^{1-t}] \tag{8.8}$$

is a joint concave function of $(\mathbf{A}_1, \mathbf{A}_2)$. *That is,*

$$\lambda\,\mathrm{tr}[\mathbf{A}_1^t\mathbf{K}^*\mathbf{A}_1^{1-t}\mathbf{K}] + (1-\lambda)\,\mathrm{tr}[\mathbf{A}_2^t\mathbf{K}^*\mathbf{A}_2^{1-t}\mathbf{K}] \le \mathrm{tr}[\mathbf{C}^t\mathbf{K}^*\mathbf{C}^{1-t}\mathbf{K}] \tag{8.9}$$

where $\mathbf{C} = \lambda\mathbf{A}_1 + (1-\lambda)\mathbf{A}_2$.

Proof. We follow a simple proof provided by Ruskai [135] below.

We first observe that

$$\langle\phi, \mathbf{C}\phi\rangle = 0 \Rightarrow \langle\phi, \mathbf{A}_1\phi\rangle = \langle\phi, \mathbf{A}_2\phi\rangle = 0,$$

so that (8.9) holds trivially for $\phi \in \ker(\mathbf{C})$. Hence, it suffices to prove the inequality (8.9) on $\ker^\perp(\mathbf{C})$ so that we can assume without loss of generality that $\mathbf{C}$ is invertible and that $\mathbf{C}^{-1}$ is bounded.

We first consider the finite-dimensional case below:

(A) Finite-dimensional case

Define $\mathbf{M} = \mathbf{C}^{(1-t)/2}\mathbf{K}\mathbf{C}^{t/2}$ and

$$f_k(t) = \mathrm{tr}[\mathbf{A}_k^t\mathbf{C}^{-t/2}\mathbf{M}^*\mathbf{C}^{-(1-t)/2}\mathbf{A}_k^{1-t}\mathbf{C}^{-(1-t)/2}\mathbf{M}\mathbf{C}^{-t/2}], \quad k = 1, 2. \tag{8.10}$$

Then (8.9) is equivalent to

$$f(t) \equiv \lambda_1 f_1(t) + (1-\lambda)f_2(t) \le \mathrm{tr}[\mathbf{M}^*\mathbf{M}]. \tag{8.11}$$

Observe that the functions $f_k(z)$ (with t replaced by $z = x + \iota y$, $x, y \in \mathbb{R}$) above can be analytically continued to the strip $0 \le \mathbb{R}(z) \le 1$. The next step is to show that each $f_k(z)$ is bounded on this strip. To do so, we first note the polar decomposition theorem (Theorem 1.8.11) implies that one can write any operator as $\mathbf{X} = \mathbf{V}|\mathbf{X}|$ with $\mathbf{V}$ unitary and $|\mathbf{X}| = \sqrt{\mathbf{X}^*\mathbf{X}}$ so that $|\mathbf{X}| = \mathbf{V}^*\mathbf{X}$. Using this in (8.10) (with t replaced by $z = x + \iota y$) yields

$$|f_k(z)| \le \|\mathbf{A}_k\|\|\mathbf{C}^{-1}\|\,\mathrm{tr}[|\mathbf{M}^*\mathbf{V}^*\mathbf{M}|] \le \|\mathbf{A}_k\|\|\mathbf{C}^{-1}\|\|\mathbf{M}\|\,\mathrm{tr}[|\mathbf{M}|]. \tag{8.12}$$

Although the unitary $\mathbf{V}$ may depend on z, the last bound on the right is independent of z. By the maximum modulus principle (see Rudin [133]), $|f_k(z)|$ is bounded by its supremum on the boundary of this strip, i. e., for $z = 0 + \iota y$ or $z = 1 + \iota y$. Now,

$$\begin{aligned} f_k(0 + \iota y) = \mathrm{tr}\big[&(\mathbf{A}_k^{\iota y/2}\mathbf{C}^{-\iota y/2}\mathbf{M}^*\mathbf{C}^{\iota y/2}\mathbf{C}^{-1/2}\mathbf{A}_k^{1/2}) \\ &\times (\mathbf{A}_k^{-\iota y}\mathbf{A}_k^{1/2}\mathbf{C}^{-1/2}\mathbf{C}^{\iota y/2}\mathbf{M}\mathbf{C}^{-\iota y/2}\mathbf{A}_k^{\iota y/2})\big]. \end{aligned} \tag{8.13}$$

Note that (8.13) has the form $\mathrm{tr}[\mathbf{X}^*\mathbf{Y}]$, which is bounded above by

$$\sqrt{\operatorname{tr}[\mathbf{X}^*\mathbf{X}]}\sqrt{\operatorname{tr}[\mathbf{Y}^*\mathbf{Y}]} \le \frac{1}{2}(\operatorname{tr}[\mathbf{X}^*\mathbf{X}] + \operatorname{tr}[\mathbf{Y}^*\mathbf{Y}])$$

and, therefore,

$$\begin{aligned}|f_k(0+\imath y)| \le{}& \frac{1}{2}\operatorname{tr}[\mathbf{M}^*\mathbf{C}^{\imath y/2}\mathbf{C}^{-1/2}\mathbf{A}_k\mathbf{C}^{-1/2}\mathbf{C}^{-\imath y/2}\mathbf{M}] \\ &+ \frac{1}{2}\operatorname{tr}[\mathbf{M}^*\mathbf{C}^{-\imath y/2}\mathbf{C}^{-1/2}\mathbf{A}_k\mathbf{C}^{-1/2}\mathbf{C}^{\imath y/2}\mathbf{M}], \quad k = 1,2.\end{aligned} \tag{8.14}$$

Thus,

$$\begin{aligned}|f(0+\imath y)| &\le \lambda|f_1(0+\imath y)| + (1-\lambda)|f_2(0+\imath y)| \\ &\le \frac{1}{2}\operatorname{tr}[\mathbf{M}^*\mathbf{C}^{\imath y/2}\mathbf{C}^{-1/2}(\lambda\mathbf{A}_1 + (1-\lambda)\mathbf{A}_2)\mathbf{C}^{-1/2}\mathbf{C}^{-\imath y/2}\mathbf{M}] \\ &\quad + \frac{1}{2}\operatorname{tr}[\mathbf{M}^*\mathbf{C}^{-\imath y/2}\mathbf{C}^{-1/2}(\lambda\mathbf{A}_1 + (1-\lambda)\mathbf{A}_2)\mathbf{C}^{-1/2}\mathbf{C}^{\imath y/2}\mathbf{M}] \\ &= \frac{1}{2}\operatorname{tr}[\mathbf{M}^*\mathbf{C}^{\imath y/2}\mathbf{C}^{-1/2}\mathbf{M}] + \frac{1}{2}\operatorname{tr}[\mathbf{M}^*\mathbf{C}^{-\imath y/2}\mathbf{C}^{\imath y}\mathbf{M}] = \operatorname{tr}[\mathbf{M}^*\mathbf{M}]\end{aligned} \tag{8.15}$$

since $\mathbf{C}$ was defined as $\lambda\mathbf{A}_1 + (1-\lambda)\mathbf{A}_2$. One can similarly show that

$$|f(1+\imath y)| \le \operatorname{tr}[\mathbf{M}^*\mathbf{M}],$$

which implies (8.11). This establishes the concavity of $\mathbf{A} \mapsto \operatorname{tr}[\mathbf{A}^t\mathbf{K}^*\mathbf{A}^{1-t}\mathbf{K}]$. The general case then follows from the observation that

$$\begin{aligned}&\operatorname{Tr}[\mathbf{A}_1^t\mathbf{K}^*\mathbf{A}_2^{1-t}\mathbf{K}] \\ &= \operatorname{tr}\left[\begin{pmatrix}\mathbf{A}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_2\end{pmatrix}^t\begin{pmatrix}\mathbf{0} & \mathbf{K}^*\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{A}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_2\end{pmatrix}^{1-t}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{K} & \mathbf{0}\end{pmatrix}\right].\end{aligned}$$

This proves the concavity of $(\mathbf{A}_1, \mathbf{A}_2) \mapsto \operatorname{tr}[\mathbf{A}_1^t\mathbf{K}^*\mathbf{A}_2^{1-t}\mathbf{K}]$ for the finite-dimensional case.

(B) Extension to infinite dimensions

The restriction to finite-dimensional matrices was used only to ensure that $\mathbf{C}^{-1}$ is bounded on the orthogonal complement of $\ker(\mathbf{C})$ so that (8.12) gives a uniform bound for $|f_k(z)|$ on the strip $0 \le \Re(z) \le 1$. It is worth emphasizing that in this part of the proof it is enough to show that $|f_k(z)|$ satisfies some upper bound, which can be rather crude as long as it holds uniformly for all z in the infinite strip. Only for the subsequent estimate on the boundary do we need a precise bound of the form (8.15), which can be generalized to operators on infinite-dimensional spaces. Therefore, the Lieb concavity theorem can be extended to infinite dimensions in several ways. First, observe that $\mathbf{C} = \lambda_1\mathbf{A}_1 + \lambda_2\mathbf{A}_2$ ($\lambda_1 = \lambda$ and $\lambda_2 = 1-\lambda$) implies that for $0 \le q \le 1$

$$\mathbf{A}_k \le \lambda_k^{-1}\mathbf{C} \Rightarrow \mathbf{A}_k^q \le \lambda_k^{-q}\mathbf{C}^q \Rightarrow \mathbf{C}^{-q/2}\mathbf{A}_k^q\mathbf{C}^{-q/2} \le \lambda_k^{-q}\mathbf{I}$$

where the first implication uses the operator monotonicity of the map $\mathbf{A} \mapsto \mathbf{A}^q$. Then under the additional hypothesis that $\mathbf{K}$ is Hilbert–Schmidt (i. e., $\mathrm{tr}[\mathbf{K}^*\mathbf{K}] < +\infty$), one can replace (8.12) by

$$|f_k(z)| \le \lambda_k^{-1}\,\mathrm{tr}[\mathbf{M}^*\mathbf{M}] \le \lambda_k^{-1}\|\mathbf{C}\|\,\mathrm{tr}[\mathbf{K}^*\mathbf{K}].$$

Lieb uses the even weaker assumption that $\mathbf{M} = \mathbf{C}^{q/2}\mathbf{K}\mathbf{C}^{p/2}$ is Hilbert–Schmidt, and also proves concavity for the map $(\mathbf{A},\mathbf{B}) \mapsto \mathrm{tr}[\mathbf{A}^p\mathbf{K}^*\mathbf{B}^q\mathbf{K}]$ for $0 \le p+q \le 1$. For this, Ruskai [135] uses the operator concavity of the map $\mathbf{A} \mapsto \mathbf{A}^t$ for $0 \le t \le 1$ to conclude that $\mathbf{C}^{-t/2}(\lambda_1\mathbf{A}_1^t + \lambda_2\mathbf{A}_2^t)\mathbf{C}^{-t/2} \le \mathbf{I}$. This proves the lemma. □

Important properties of relative entropy are summarized in the following theorem (see Wehrl [175] and Ohya and Petz [121]). We only provide proofs for some of them, since many of these proofs follow the same line of that in the previous chapter that deals with von Neumann entropy.

Theorem 8.1.6. *The relative entropy satisfies the following properties:*

1. (Positivity) $H(\rho\|\sigma) \ge 0$, $H(\sigma\|\sigma) = 0$ *if and only if* $\rho = \sigma$.
2. (Joint convexity) $H(\lambda\rho_1 + (1-\lambda)\rho_2\|\lambda\sigma_1 + (1-\lambda)\sigma_2) \le \lambda H(\rho_1\|\sigma_1) + (1-\lambda)H(\rho_2\|\sigma_2)$ *for any* $\lambda \in [0,1]$.
3. (Additivity) $H(\rho_1 \otimes \rho_2\|\sigma_1 \otimes \sigma_2) = H(\rho_1\|\sigma_1) + H(\rho_2\|\sigma_2)$.
4. (Lower semicontinuity) *If* $\lim_{n\to+\infty}\|\rho_n - \rho\|_1 = 0$ *and* $\lim_{n\to+\infty}\|\sigma_n - \sigma\|_1 = 0$, *then* $H(\rho\|\sigma) \le \liminf_{n\to+\infty} H(\rho_n\|\sigma_n)$. *Moreover, if there exists a positive number* λ *satisfying* $\rho_n \le \lambda\sigma_n$, *then* $\lim_{n\to+\infty} H(\rho_n\|\sigma_n) = H(\rho\|\sigma)$.
5. (Lower bound) $\frac{1}{2}\|\rho - \sigma\|_1^2 \le H(\rho\|\sigma)$.
6. (Invariance under the unitary mapping) $H(\mathbf{U}\rho\mathbf{U}^*\|\mathbf{U}\sigma\mathbf{U}^*) = H(\rho\|\sigma)$, *where* $\mathbf{U}$ *is a unitary operator.*
7. (Donald's identity) *Suppose the density operator* ρ_k *occurs with probability* p_k, *yielding an average state* $\rho = \sum_k p_k\rho_k$, *and suppose* σ *is some other density operator. Then*

$$\sum_k p_k H(\rho_k\|\rho) = \sum_k p_k H(\rho_k\|\rho) + H(\rho\|\sigma)$$

Proof. To avoid repetition of proof, we only prove parts of this theorem and leave the rest to the readers to fill in the gaps.

2. (Joint convexity) To prove the joint convexity of $H(\cdot\|\cdot)$, we consider joint concavity in ρ_1 and ρ_2 of the function $\mathrm{tr}[\rho_1^t\rho_2^{1-t}]$ by using of the Lieb's concavity theorem (8.1.5) with $\mathbf{K} = \mathbf{I}$:

$$\frac{d}{dt}\,\mathrm{tr}[\rho_1^t\rho_2^{1-t}]\big|_{t=0} = \mathrm{tr}[\rho_2(\log\rho_1 - \log\rho_2)],$$

is joint concave in ρ_1 and ρ_2. Therefore, $H(\rho_1\|\rho_2)$ is jointly convex. This proves part 2.

4. (Lower semicontinuity) Since the function $\eta(x) = -x \log x$ for $x > 0$ and $= 0$ for $x = 0$ is continuous on the compact interval $[0, 1]$, it is easy to show that $\lim_{n\to+\infty} \|\rho_n - \rho\|_1 = 0$ implies that $\lim_{n\to+\infty} |\eta(\rho_n) - \eta(\rho)| = 0$, since

$$\|\mathbf{A}\|_1 = \sup_{\phi\neq 0} \frac{\langle\phi|\mathbf{A}|\phi\rangle_{\mathbb{H}}}{\langle\phi,\phi\rangle_{\mathbb{H}}}.$$

Thus, for the finite-dimensional projection operator $\mathbf{P}$ on $\mathbb{H}$, $\lim_{n\to+\infty} \|\mathbf{P}(\eta(\rho_n) - \eta(\rho))\|_1 = 0$ because of the standard inequality $\mathrm{tr}[\mathbf{PA}] \leq \mathrm{tr}[\mathbf{P}\|\mathbf{A}\|]$. On the other hand,

$$\mathrm{tr}[\mathbf{PA}] \leq \|\mathbf{P}\|_\infty \,\mathrm{tr}[\mathbf{A}] \quad \text{and} \quad \mathrm{tr}[\mathbf{A}] \leq \sup_{\mathbf{P}} \mathrm{tr}[\mathbf{PA}].$$

Therefore,

$$H(\rho) = \sup_{\mathbf{P}} \mathrm{tr}[\mathbf{P}\eta(\rho)] \leq \liminf_{n\to+\infty}\Big(\sup_{\mathbf{P}} \mathrm{tr}[\mathbf{P}\eta(\rho_n)]\Big) = \liminf_{n\to+\infty} H(\rho_n).$$

Now,

$$\begin{aligned} H(\sigma\|\rho) &= \sup_{\mathbf{P},\lambda} \mathrm{tr}\big[\mathbf{P}(\eta\big(\lambda\sigma + (1-\lambda)\rho\big) + \lambda\eta(\sigma_n))\big] \\ &\leq \liminf_{n\to+\infty}(\sup_{\mathbf{P},\lambda} \mathrm{tr}\big[\mathbf{P}(\eta\big(\lambda\sigma_n + (1-\lambda)\rho_n\big) + \lambda\eta(\sigma_n) - (1-\lambda)\eta(\rho_n))\big]\big) \\ &= \liminf_{n\to+\infty} H(\sigma_n\|\rho_n). \end{aligned}$$

This proves the lower semicontinuity of $H(\cdot\|\cdot)$.

7. The proof of Donald's identity for relative entropy is given below:

$$\begin{aligned} \sum_k p_k H(\rho_k\|\sigma) &= \sum_k p_k(\mathrm{tr}[\rho_k \log \rho_k] - \mathrm{tr}[\rho_k \log \sigma]) \\ &= \sum_k p_k(\mathrm{tr}[\rho_k \log \rho_k] - \mathrm{tr}[\rho_k \log \rho] + \mathrm{tr}[\rho_k \log \rho] - \mathrm{tr}[\rho_k \log \sigma]) \\ &= \sum_k p_k(\mathrm{tr}[\rho_k \log \rho_k] - \mathrm{tr}[\rho_k \log \rho]) - \mathrm{tr}[\rho \log \rho - \rho \log \sigma] \\ &= \sum_k p_k H(\rho_k\|\rho) + H(\rho\|\sigma). \end{aligned}$$

This proves Donald's identity of $H(\cdot\|\cdot)$. □

Corollary 8.1.7. *Let ρ and σ be two quantum states on $\mathbb{H}$, then*

$$H\big(\lambda\rho + (1-\lambda)\sigma\big) \geq \lambda H(\rho) + (1-\lambda)H(\sigma) + \frac{\lambda(1-\lambda)}{2}\|\rho - \sigma\|_1, \quad \forall \lambda \in [0, 1].$$

Proof. We apply part 5 of Theorem 8.1.6 to each of the following two quantities and obtain

$$\lambda H(\rho\|\lambda\rho + (1-\lambda)\sigma) + (1-\lambda)H(\sigma\|\lambda\rho + (1-\lambda)\sigma)$$
$$\geq \frac{1}{2}\lambda\|(1-\lambda)(\rho-\sigma)\|_1^2 + \frac{1}{2}(1-\lambda)\|\lambda(\rho-\sigma)\|_1^2 = \frac{\lambda(1-\lambda)}{2}\|\rho-\sigma\|_1^2.$$

The above inequality implies that

$$H(\lambda\rho + (1-\lambda)\sigma) \geq \lambda H(\rho) + (1-\lambda)H(\sigma) + \frac{\lambda(1-\lambda)}{2}\|\rho-\sigma\|_1^2.$$

This proves the corollary. □

Note if $\rho = \sigma \in \mathcal{S}(\mathbb{H})$ that $H(\rho\|\sigma) = 0$. Despite this fact the relative entropy is not a metric, because it is nonsymmetric, i. e., $H(\rho\|\sigma) \neq H(\sigma\|\rho)$ in general. However, it is possible to introduce the notion of convergence of a sequence of states $(\rho_n)_{n=1}^{+\infty}$ to a particular state ρ as follows.

Definition 8.1.8 (H-convergence). A sequence $(\rho_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H})$ is said to H-converges to $\rho \in \mathcal{S}(\mathbb{H})$ if

$$\lim_{n\to+\infty} H(\rho_n\|\rho) = 0. \tag{8.16}$$

In this case, we write $(H)\lim_{n\to+\infty}\rho_n = \rho$.

Part 5 of Theorem 8.1.6 implies that the metric induced by the relative entropy function $H(\cdot\|\cdot) : \mathcal{S}(\mathbb{H}) \times \mathcal{S}(\mathbb{H}) \to [0, +\infty]$ (or simply the H-metrics) is weaker than the metric induced by the trace-norm $\|\cdot\|_1$.

Proposition 8.1.9. *A sequence of quantum states $(\rho_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H})$ converges to $\rho \in \mathcal{S}(\mathbb{H})$ in trace-norm $\|\cdot\|_1$ if and only if it converges in H-topology defined in Definition* 8.1.8. *That is,*

$$(H)\lim_{n\to\infty}\rho_n = \rho \quad \textit{if and only if} \ \lim_{n\to\infty}\|\rho_n - \rho\|_1 = 0. \tag{8.17}$$

Proof. (⇒) It is clear from part 5 of Theorem 8.1.6 that H-convergence implies convergence in trace-norm $\|\cdot\|_1$.

(⇐) Assume that $\lim_{n\to+\infty}\|\rho_n - \rho\|_1 = 0$ and that $\ker(\rho_n)^\perp \subseteq \ker(\rho)^\perp$ for all n. Then by the definition of $H(\rho_n\|\rho)$ (Definition 8.1.3), we have

$$\begin{aligned} H(\rho_n\|\rho) &= \sum_i \langle i|\rho_n \log\rho_n - \rho_n\log\rho + \rho - \rho_n|i\rangle_{\mathbb{H}} \\ &= \mathrm{tr}[\rho_n(\log\rho_n - \log\rho)] + \mathrm{tr}[\rho - \rho_n], \\ &\to 0, \end{aligned}$$

since $\|\rho_n - \rho\|_1 \to 0$ implies $\mathrm{tr}[\rho_n(\log\rho_n - \log\rho)] \leq \|\rho_n\|_\infty\|\log\rho_n - \log\rho\|_1 \to 0$ as $n \to +\infty$. This proves the proposition. □

8.2 $H(\cdot\|\cdot)$ extended to $\mathfrak{T}_+(\mathbb{H}) \times \mathfrak{T}_+(\mathbb{H})$

By the same token in which von Neumann entropy $H(\cdot)$ is extended to $\mathfrak{T}_+(\mathbb{H})$, the extended relative entropy $H(\cdot\|\cdot) : \mathfrak{T}_+(\mathbb{H})\times\mathfrak{T}_+(\mathbb{H}) \to [-\infty, +\infty]$ can be defined as follows.

Definition 8.2.1. The relative entropy of operators $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$ is defined as follows:

$$H(\mathbf{A}\|\mathbf{B}) = \begin{cases} \sum_{i=1}^{\infty} \langle e_i|(\mathbf{A}\log\mathbf{A} - \mathbf{A}\log\mathbf{B} + \mathbf{B} - \mathbf{A})|e_i\rangle_{\mathbb{H}}, & \ker^{\perp}(\mathbf{A}) \subseteq \ker^{\perp}(\mathbf{B}), \\ +\infty, & \text{otherwise}, \end{cases}$$

where $\{e_i\}_{i=1}^{+\infty}$ is an orthonormal basis of eigenvectors of the operator $\mathbf{A}$ and the series consists of nonnegative terms.

Note that the extended relative entropy $H(\cdot\|\cdot)$ defined above reduces to the relative entropy defined on $\mathcal{S}(\mathbb{H}) \times \mathcal{S}(\mathbb{H})$ because $\operatorname{tr}[\mathbf{B}] = \operatorname{tr}[\mathbf{A}] = 1$ when $\mathbf{A}, \mathbf{B} \in \mathcal{S}(\mathbb{H})$.

Lemma 8.2.2. *Let ρ and σ be quantum states in $\mathcal{S}(\mathbb{H})$, and let $\mathbf{C}$ be an operator in $\mathfrak{T}_+(\mathbb{H})$. Then*

$$H(\lambda\rho + (1-\lambda)\sigma\|\mathbf{C}) \geq \lambda H(\rho\|\mathbf{C}) + (1-\lambda)H(\sigma\|\mathbf{C}) - h_2(\lambda), \quad \lambda \in [0,1],$$

where $h_2(\lambda) = \eta(\lambda) + \eta(1-\lambda)$.

Proof. Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be an increasing sequence of finite-rank projection operators strongly converging to the identity operator $\mathbf{I}_{\mathbb{H}}$. Then $\mathbf{A}_n = \mathbf{P}_n\rho\mathbf{P}_n$, $\mathbf{B}_n = \mathbf{P}_n\sigma\mathbf{P}_n$ and $\mathbf{C}_n = \mathbf{P}_n\mathbf{C}\mathbf{P}_n$ are finite-rank operators for each n, and hence,

$$\begin{aligned}
&H(\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n\|\mathbf{C}_n) \\
&\quad = \operatorname{tr}\big[(\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n)\log(\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n)\big] \\
&\qquad - \operatorname{tr}\big[(\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n)\log\mathbf{C}_n\big] \\
&\qquad + \operatorname{tr}[\mathbf{C}_n] - \operatorname{tr}\big[\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n\big] \\
&\quad = \operatorname{tr}\big[(\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n(-\log(\mathbf{C}_n))\big] - \operatorname{tr}\big[\eta(\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n)\big] \\
&\qquad + \operatorname{tr}[\mathbf{C}_n] - \operatorname{tr}\big[\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n\big] \\
&\quad \geq \lambda\operatorname{tr}\big[\mathbf{A}_n(-\log\mathbf{C}_n)\big] + (1-\lambda)\operatorname{tr}\big[\mathbf{B}_n(-\log\mathbf{C}_n)\big] + \operatorname{tr}[\mathbf{C}_n] \\
&\qquad - \lambda\operatorname{tr}[\mathbf{A}_n] - (1-\lambda)\operatorname{tr}[\mathbf{B}_n] - \lambda\operatorname{tr}[\eta(\mathbf{A}_n)] - (1-\lambda)\operatorname{tr}[\eta(\mathbf{B}_n)] \\
&\qquad - \eta\big(\operatorname{tr}[\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n]\big) + \lambda\eta\big(\operatorname{tr}[\mathbf{A}_n]\big) + (1-\lambda)\eta\big(\operatorname{tr}[\mathbf{B}_n]\big) \\
&\qquad - x_n h_2\big(x_n^{-1}\lambda\operatorname{tr}[\mathbf{A}_n]\big) \\
&\quad = \lambda H(\mathbf{A}_n\|\mathbf{C}_n) + (1-\lambda)H(\mathbf{B}_n\|\mathbf{C}_n) - \eta\big(\operatorname{tr}[\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n]\big) \\
&\qquad + \lambda\eta\big(\operatorname{tr}[\mathbf{A}_n]\big) + (1-\lambda)\eta\big(\operatorname{tr}[\mathbf{B}_n]\big) - x_n h_2\big(x_n^{-1}\lambda\operatorname{tr}[\mathbf{A}_n]\big),
\end{aligned}$$

where $x_n = \operatorname{tr}[\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n]$ and where we used the inequality

$$H(\lambda\mathbf{A}_n + (1-\lambda)\mathbf{B}_n) \leq \lambda H(\mathbf{A}_n) + (1-\lambda)H(\mathbf{B}_n) + x_n h_2\big(x_n^{-1}\lambda\operatorname{tr}(\mathbf{A}_n)\big),$$

following (7.43) and (8.18). By Lemma 8.2.5 and passing to the limit, we have proved the lemma. □

Lemma 8.2.3. *Let $(\mathbf{A}_n)_{n=1}^{+\infty}$ be a sequence of operators in $\mathfrak{T}_+(\mathbb{H})$ converging in the trace-norm $\|\cdot\|_1$ to an operator $\mathbf{A}_0$ such that $\mathbf{A}_n \le \mathbf{A}_0$ for all n. Then*

$$\lim_{n\to+\infty} H(\mathbf{A}_n\|\mathbf{B}) = H(\mathbf{A}_0\|\mathbf{B}),$$

for any operator $\mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$.

Proof. Without loss of generality, we can assume that $\mathbf{A}_0 \in \mathcal{S}(\mathbb{H})$ is a quantum state. It can be represented in the form

$$\mathbf{A}_0 = \lambda_n\rho_n + (1-\lambda_n)\sigma_n,$$

where

$$\lambda_n = \mathrm{tr}[\mathbf{A}_n], \quad \rho_n = \frac{\mathbf{A}_n}{\mathrm{tr}]\mathbf{A}_n]}, \quad \sigma_n = \frac{\mathbf{A}_0 - \mathbf{A}_n}{1-\lambda_n}.$$

By Lemma 8.2.2 and nonnegativity of relative entropy, we have

$$\begin{aligned} H(\mathbf{A}_0\|\mathbf{B}) &\ge \lambda_n H(\rho_n\|\mathbf{B}) + (1-\lambda_n)H(\sigma_n\|\mathbf{B}) - h_2(\lambda_n) \\ &\ge H(\mathbf{A}_n\|\lambda_n\mathbf{B}) - h_2(\lambda_n) \\ &= H(\mathbf{A}_n\|\mathbf{B}) - \mathrm{tr}(\mathbf{B})(1-\lambda_n) - \lambda_n\log(\lambda_n) - h_2(\lambda_n). \end{aligned}$$

Hence, $\lim_{n\to+\infty} H(\mathbf{A}_n\|\mathbf{B}) \le H(\mathbf{A}_0\|\mathbf{B})$. By lower semicontinuity of $H(\cdot\|\cdot)$, this proves the lemma. □

Lemma 8.2.4. *Let $\mathbf{P}$ be an orthogonal projection on $\mathbb{H}$ and $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$. Then:*
1. $H(\mathbf{PAP}) \le H(\mathbf{A})$.
2. $H(\mathbf{PAP}^*\|(\mathbf{I}-\mathbf{P})\mathbf{A}(\mathbf{I}-\mathbf{P})) + H(\mathbf{PBP}^*\|(\mathbf{I}-\mathbf{P})\mathbf{B}(\mathbf{I}-\mathbf{P})) \le H(\mathbf{A}\|\mathbf{B})$.

Proof. 1. Since $\mathbf{PAP} \le \mathbf{A}$ and $H(\cdot)$ is nondecreasing, $H(\mathbf{PAP}) \le H(\mathbf{A})$.

2. Note that $\mathbf{U} = 2\mathbf{P}_\mathbf{I}$ is unitary and that

$$\mathbf{PAP} + (\mathbf{I}-\mathbf{P})\mathbf{A}(\mathbf{I}-\mathbf{P}) = \frac{1}{2}(\mathbf{A} + \mathbf{U}^*\mathbf{AU}).$$

By the joint convexity of the relative entropy $H(\cdot\|\cdot)$, we have

$$\begin{aligned} &H(\mathbf{PAP}\|\mathbf{PBP}) + H((\mathbf{I}-\mathbf{P})\mathbf{A}(\mathbf{I}-\mathbf{P})\|(\mathbf{I}-\mathbf{P})\mathbf{B}(\mathbf{I}-\mathbf{P})) \\ &\quad = H(\mathbf{PAP} + (\mathbf{I}-\mathbf{P})\mathbf{A}(\mathbf{I}-\mathbf{P})\|\mathbf{PBP} + (\mathbf{I}-\mathbf{P})\mathbf{B}(\mathbf{I}-\mathbf{P})) \\ &\quad \le \frac{1}{2}H(\mathbf{A}\|\mathbf{B}) + \frac{1}{2}H(\mathbf{U}^*\mathbf{AU}\|\mathbf{U}^*\mathbf{BU}) = H(\mathbf{A}\|\mathbf{B}). \end{aligned}$$

This proves the lemma. □

The following result (due originally to Lindblad [108]) provides a tool for approximation of infinite-dimensional quantum entropies.

Lemma 8.2.5. *Let* $(\mathbf{P}_n)_{n=1}^{+\infty}$ *be an arbitrary sequence of finite-dimensional projectors monotonously increasing to the unit operator* $\mathbf{I}_{\mathbb{H}}$ *and let* $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$. *Then the sequences* $(H(\mathbf{P}_n\mathbf{A}\mathbf{P}_n))_{n=1}^{+\infty}$ *and* $(H(\mathbf{P}_n\mathbf{A}\mathbf{P}_n\|\mathbf{P}_n\mathbf{B}\mathbf{P}_n))_{n=1}^{+\infty}$ *are monotonously increasing and have the limits*

$$H(\mathbf{A}) = \lim_{n\to+\infty} H(\mathbf{P}_n\mathbf{A}\mathbf{P}_n) \quad \textit{and} \quad H(\mathbf{A}\|\mathbf{B}) = \lim_{n\to+\infty} H(\mathbf{P}_n\mathbf{A}\mathbf{P}_n\|\mathbf{P}_n\mathbf{B}\mathbf{P}_n)$$

in the range $[0, +\infty]$ *independent of the choice of the sequence* $(\mathbf{P}_n)_{n=1}^{+\infty}$.

Proof. We follow the proof of Lindblad [108]. First, note that $\mathbf{A}_n = \mathbf{P}_n\mathbf{A}\mathbf{P}_n \in \mathfrak{T}_+(\mathbb{H})$ converges to $\mathbf{A}$ uniformly when $n \to +\infty$. We can write

$$\mathbf{A} - \mathbf{A}_n = \mathbf{A}_n^+ - \mathbf{A}_n^- = (\mathbf{P}_n\mathbf{A}\mathbf{P}_n)^+ - (\mathbf{P}_n\mathbf{A}\mathbf{P}_n)^-,$$

where $\mathbf{A}_n^+, \mathbf{A}_n^- \in \mathfrak{T}_+(\mathbb{H})$ for each $n \in \mathbb{N}$ and $\mathbf{A}_n^+\mathbf{A}_n^- = \mathbf{0}$. Obviously, $\lim_{n\to+\infty} \mathrm{tr}[\mathbf{A} - \mathbf{A}_n] = 0$. It follows from the definitions of extended von Neumann entropy $H(\cdot)$ and extended relative entropy $H(\cdot\|\cdot)$ that

$$\lim_{n\to+\infty} H(\mathbf{P}_n\mathbf{A}\mathbf{P}_n) = H(\mathbf{A})$$

and

$$\lim_{n\to+\infty} H(\mathbf{P}_n\mathbf{A}\mathbf{P}_n\|\mathbf{P}_n\mathbf{B}\mathbf{P}_n) = H(\mathbf{A}\|\mathbf{B}).$$

This proves the lemma. □

With the help of Lemma 8.2.5, the above definitions of entropy and relative entropy can also be defined for $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$ as follows:

$$H(\mathbf{A}) = \lim_{n\to+\infty} H(\mathbf{P}_n\mathbf{A}\mathbf{P}_n) \tag{8.18}$$

and

$$H(\mathbf{A}\|\mathbf{B}) = \lim_{n\to+\infty} H(\mathbf{P}_n\mathbf{A}\mathbf{P}_n\|\mathbf{P}_n\mathbf{B}\mathbf{P}_n), \tag{8.19}$$

where $(\mathbf{P}_n)_{n=1}^{+\infty}$ be an arbitrary sequence of finite-dimensional projectors monotonously increasing to the unit operator $\mathbf{I}_{\mathbb{H}}$.

The proof of the following lemma follows exactly from part 4 of Theorem 8.1.6 with very minor modification.

Lemma 8.2.6. *Let* $(\mathbf{A}_n)_{n=1}^{+\infty} \subset \mathfrak{T}_+(\mathbb{H})$ *and* $(\mathbf{B}_n)_{n=1}^{+\infty} \subset \mathfrak{T}_+(\mathbb{H})$ *be such that*

$$\lim_{n\to+\infty} \|\mathbf{A}_n - \mathbf{A}\|_1 = 0 \quad \textit{and} \quad \lim_{n\to+\infty} \|\mathbf{B}_n - \mathbf{B}\|_1 = 0 \quad \textit{for some } \mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H}).$$

Then

$$H(\mathbf{A}\|\mathbf{B}) \le \liminf_{n\to+\infty} H(\mathbf{A}_n\|\mathbf{B}_n) \tag{8.20}$$

The first inequality of the following lemma follows from strong subadditivity as formulated in Lemma 7.2.10. The second statement is trivial.

Lemma 8.2.7. *Let* $\mathbb{H} = \mathbb{H}_1 \otimes \mathbb{H}_2$ *and put* $\mathbf{A}_1 := \mathrm{tr}_2[\mathbf{A}]$ *and* $\mathbf{B}_1 := \mathrm{tr}_2[\mathbf{B}]$ *for all* $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_+(\mathbb{H})$, *where* $\mathrm{tr}_2[\cdot] = \mathrm{tr}_{\mathbb{H}_2}[\cdot]$. *Then*

$$H(\mathbf{A}_1\|\mathbf{B}_1) \le H(\mathbf{A}\|\mathbf{B}). \tag{8.21}$$

Furthermore, if $\mathbf{A}_1, \mathbf{B}_1 \in \mathfrak{T}_+(\mathbb{H}_1)$, $\mathbf{A}_2 \in \mathfrak{T}_+(\mathbb{H}_2)$, *and* $\mathrm{tr}[\mathbf{A}_2] = 1$, *then*

$$H(\mathbf{A}_1 \otimes \mathbf{A}_2\|\mathbf{B}_1 \otimes \mathbf{A}_2) = H(\mathbf{A}_1\|\mathbf{B}_1). \tag{8.22}$$

8.3 Quantum conditional entropy

The conditional quantum entropy is an entropy measure used in quantum information theory. It is a generalization of the conditional entropy of classical information theory. For a bipartite state ρ_{AB}, the conditional entropy is written as $H(A|B)_\rho$. The quantum conditional entropy was defined in terms of a conditional density operator $\rho_{A|B}$ by Cerf and Adami [19, 20] who showed that quantum conditional entropies can be negative, something that is forbidden in classical physics. The negativity of quantum conditional entropy is a sufficient criterion for quantum nonseparability (see Chapter 10 for separability of quantum states).

8.3.1 Definitions

Consider a bipartite composite quantum systems AB represented by the Hilbert space $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$, where $\mathbb{H}_A$ and $\mathbb{H}_B$ for the component system A and B, respectively.

Assume first that $\dim(\mathbb{H}_A) < +\infty$ and $\dim(\mathbb{H}_B) < +\infty$. Let $\rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$ be a joint state of composite system AB, and $\rho_A = \mathrm{tr}_B[\rho_{AB}]$ and $\rho_B = \mathrm{tr}_A[\rho_{AB}]$ are corresponding partial states.

Definition 8.3.1 (Finite-dimensional case). Let $\rho = \rho_{AB}$ be a quantum state in the composite system AB of the finite-dimensional Hilbert space $\mathbb{H}_{AB}$. The conditional entropy $H(A|B)_\rho$ of A given B under the joint quantum state $\rho = \rho_{AB}$ is defined by

$$H(A|B)_\rho = H(\rho_{AB}) - H(\rho_B), \tag{8.23}$$

where $H(\cdot) = \mathrm{tr}[\eta(\cdot)]$ is the von Neumann entropy defined in (7.1.1).

The conditional entropy (8.23) is well-defined in finite-dimensional Hilbert spaces and takes its values in $]-\infty, +\infty[$. However, in the infinite-dimensional case, uncertainty may occur in the sense that we may have the situation that $(+\infty) - (+\infty)$. Therefore, we introduce the following definition for infinite- dimensional $\mathbb{H}_A$ and $\mathbb{H}_B$.

Definition 8.3.2 (Infinite-dimensional case). Let $\mathbb{H}_A$ and $\mathbb{H}_B$ be arbitrary complex separable Hilbert spaces with $\rho = \rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$ and $H(\rho_A) < +\infty$. Then we define a conditional entropy of A given B under ρ_{AB} as

$$H(A|B)_\rho = H(\rho_A) - H(\rho_{AB} \| \rho_A \otimes \rho_B), \tag{8.24}$$

where $H(\cdot\|\cdot)$ is the relative entropy defined in Definition 8.2.1.

Obviously, the conditional entropy defined by relation (8.24) is equivalent to (8.23) on the finite rank states, since

$$H(\rho_{AB} \| \rho_A \otimes \rho_B) = H(\rho_A) + H(\rho_B) - H(\rho_{AB}).$$

Due to the condition $H(\rho_A) < +\infty$, Definition 8.3.2 is well-defined. Inequalities $H(\rho_A) < +\infty$, $H(\rho_{AB} \| \rho_A \otimes \rho_B) \geq 0$ imply that

$$-\infty \leq H(A|B)_\rho \leq H(\rho_A) < +\infty. \tag{8.25}$$

In the above and in what follows, we use the shorthand notation $H(A) = H(\rho_A)$, $H(B) = H(\rho_B)$, etc., whenever there is no danger of ambiguity.

8.3.2 Properties of conditional entropy

The following supplemental results regarding operators on a composite and its component Hilbert spaces will be needed in order to establish properties of conditional entropy.

The presentation of this subsection is largely based on the results obtained by Kuznetsova [102].

Lemma 8.3.3. *If sequences of projectors $(\mathbf{P}_k^A)_{k=1}^{+\infty}$ and $(\mathbf{P}_k^B)_{k=1}^{+\infty}$ on $\mathbb{H}_A$ and $\mathbb{H}_B$, converge in the operator norm $\|\cdot\|_\infty$ to the identity operators $\mathbf{I}_A$ and $\mathbf{I}_B$, respectively, then the sequence $(\mathbf{P}_k^A \otimes \mathbf{P}_k^B)_{k=1}^{+\infty}$ converges in the operator norm to the identity operator $\mathbf{I}_{AB}$ on $\mathbb{H}_{AB}$.*

Proof. We consider the following inequalities obtained via the triangular inequality of the operator norm $\|\cdot\|_\infty$:

$$\begin{aligned}&\|\mathbf{P}_k^A \otimes \mathbf{P}_k^B - \mathbf{I}_{AB}\|_\infty \\ &\quad \le \|\mathbf{P}_k^A \otimes \mathbf{P}_k^B - \mathbf{P}_k^A \otimes \mathbf{I}_B\|_\infty + \|\mathbf{P}_k^A \otimes \mathbf{I}_B - \mathbf{I}_{AB}\|_\infty \\ &\quad \le \|\mathbf{P}_k^A \otimes (\mathbf{P}_k^B - \mathbf{I}_B)\|_\infty + \|(\mathbf{P}_k^A - \mathbf{I}_A) \otimes \mathbf{I}_B\|_\infty \\ &\quad \le \|\mathbf{P}_k^A\|_\infty \|\mathbf{P}_k^B - \mathbf{I}_B\|_\infty + \|\mathbf{P}_k^A - \mathbf{I}_A\|_\infty \|\mathbf{I}_B\|_\infty,\end{aligned}$$

where we have used the fact that $\|\mathbf{A} \otimes \mathbf{B}\|_\infty \le \|\mathbf{A}\|_\infty \|\mathbf{B}\|_\infty$ in the above for operators $\mathbf{A}$ and $\mathbf{B}$ on Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. Therefore, $\lim_{k\to+\infty} \|\mathbf{P}_k^A \otimes \mathbf{P}_k^B - \mathbf{I}_{AB}\|_\infty = 0$, since $\lim_{k\to+\infty} \|\mathbf{P}_k^A - \mathbf{I}_A\|_\infty = 0$ and $\lim_{k\to+\infty} \|\mathbf{P}_k^B - \mathbf{I}_B\|_\infty = 0$. This proves the lemma. □

Lemma 8.3.4. *Let $(\mathbf{T}_k)_{k=1}^{+\infty} \subset \mathfrak{T}(\mathbb{H}_B)$ be a sequence of trace-class operators that converges in the* weak**-topology to $\mathbf{T} \in \mathfrak{T}(\mathbb{H}_B)$. Then the sequence $(\mathbf{I}_A \otimes \mathbf{T}_k)_{k=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H}_{AB})$ converges to $\mathbf{I}_A \otimes \mathbf{T}$ in the weak operator topology.*

Proof. For each $\psi \in \mathbb{H}_{AB}$, we find that

$$\langle \psi | \mathbf{I}_A \otimes \mathbf{T}_k | \psi \rangle_{AB} = \mathrm{tr}[\mathbf{T}_k \mathbf{K}_\psi^B],$$

where $\mathbf{K}_\psi^B = \mathrm{tr}_A[|\psi\rangle_{AB}\langle\psi|]$ is the reduced operator. Since $\mathbf{K}_\psi^B$ is trace class (and thus compact), we have

$$\lim_{k\to+\infty} \langle \psi | \mathbf{I}_A \otimes \mathbf{T}_k | \psi \rangle_{AB} = \lim_{k\to+\infty} \mathrm{tr}[\mathbf{T}_k \mathbf{K}_\psi^B] = \mathrm{tr}[\mathbf{T}\mathbf{K}_\psi^B] = \langle \psi | \mathbf{I}_A \otimes \mathbf{T} | \psi \rangle_{AB}.$$

That is, the sequence $(\mathbf{I}_A \otimes \mathbf{T}_k)_{k=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H}_{AB})$ converges to $\mathbf{I}_A \otimes \mathbf{T}$ in the weak* operator topology. This proves the lemma. □

For $\rho = \rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$, denote $\rho_A = \mathrm{tr}_B[\rho_{AB}]$ and $\rho_B = \mathrm{tr}_A[\rho_{AB}]$, where $\mathrm{tr}_A[\cdots]$ and $\mathrm{tr}_B[\cdots]$ are partial traces of $[\cdots]$ over the Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively.

Lemma 8.3.5. *Consider arbitrary increasing sequences of the finite rank projectors $(\mathbf{P}_n^A)_{n=1}^{+\infty}$ and $(\mathbf{P}_k^B)_{k=1}^{+\infty}$, converging in operator norm to the operator $\mathbf{I}_A$ and $\mathbf{I}_B$, respectively, and the density operator*

$$\rho_{AB}^{nk} = \lambda_{nk}^{-1}((\mathbf{P}_n^A \otimes \mathbf{P}_k^B)\rho_{AB}(\mathbf{P}_n^A \otimes \mathbf{P}_k^B)), \quad \lambda_{nk} = \mathrm{tr}[(\mathbf{P}_n^A \otimes \mathbf{P}_k^B)\rho_{AB}],$$

with partial states $\rho_A^{nk} = \mathrm{tr}_B[\rho_{AB}^{nk}]$ and $\rho_B^{nk} = \mathrm{tr}_A[\rho_{AB}^{nk}]$. Then

$$\lim_{n,k\to\infty} H(A_{nk}|B_{nk})_\rho = H(A|B)_\rho,$$

where $H(A_{nk}|B_{nk})_\rho = H(\rho_A^{nk}) - H(\rho_{AB}^{nk}\|\rho_A^{nk} \otimes \rho_B^{nk})$ and $H(A|B)_\rho = H(\rho_A) - H(\rho_{AB}\|\rho_A \otimes \rho_B)$.

Proof. By using the lower semicontinuity of the von Neumann entropy $H(\cdot)$, we obtain

$$\begin{aligned}
&\liminf_{n,k\to+\infty} H(\lambda_{nk}\rho_A^{nk})\\
&\quad= \liminf_{n,k\to+\infty} H((\mathbf{P}_n^A\otimes\mathbf{P}_k^B)\rho_A(\mathbf{P}_n^A\otimes\mathbf{P}_k^B))\\
&\quad\geq H\Big(\lim_{n,k\to+\infty}(\mathbf{P}_n^A\otimes\mathbf{P}_k^B)\rho_A(\mathbf{P}_n^A\otimes\mathbf{P}_k^B)\Big)\\
&\quad= H\Big(\lim_{n\to+\infty}(\mathbf{P}_n^A\otimes\mathbf{I}_B)\rho_A(\mathbf{P}_n^A\otimes\mathbf{I}_B)\Big)\quad\text{by Lemmas 8.3.3 and 8.3.4}\\
&\quad= H\Big(\lim_{n\to+\infty}\mathbf{P}_n^A\rho_A\mathbf{P}_n^A\Big)\quad\text{by Lemmas 8.3.3 and 8.3.4}\\
&\quad= H(\rho_A).
\end{aligned}$$

On the other hand, Lemma 8.2.5 implies

$$\begin{aligned}
H(\lambda_{nk}\rho_A^{nk}) &= H(\mathbf{P}_n^A\,\mathrm{tr}_B[(\mathbf{I}_A\otimes\mathbf{P}_k^B)\rho_{AB}(\mathbf{I}_A\otimes\mathbf{P}_k^B)]\mathbf{P}_n^A)\\
&\leq H(\mathrm{tr}_B[(\mathbf{I}_A\otimes\mathbf{P}_k^B)\rho_{AB}(\mathbf{I}_A\otimes\mathbf{P}_k^B)]).
\end{aligned}$$

Furthermore, by the dominated convergence theorem for entropy (see Lemma 7.2.11), we have

$$\lim_{n,k\to\infty} H(\mathrm{tr}_B[(\mathbf{I}_A\otimes\mathbf{P}_k^B)\rho_{AB}(\mathbf{I}_A\otimes\mathbf{P}_k^B)]) = H(\rho_A),$$

since $\mathrm{tr}_B[(\mathbf{I}_A\otimes\mathbf{P}_k^B)\rho_{AB}(\mathbf{I}_A\otimes\mathbf{P}_k^B)]\leq\rho_A$ and $H(\rho_A) < +\infty$. Thus, by the theorem on the limit of the intermediate sequence, we obtain

$$\lim_{n,k\to+\infty} H(\lambda_{nk}\rho_A^{nk}) = H(\rho_A),\quad\text{hence}\quad \lim_{n,k\to+\infty} H(\rho_A^{nk}) = H(\rho_A).$$

Then we prove that

$$\lim_{n,k\to+\infty} H(\rho_{AB}^{nk}\|\rho_A^{nk}\otimes\rho_B^{nk}) = H(\rho_{AB}\|\rho_A\otimes\rho_B).$$

Consider the following values:

$$\begin{aligned}
H_{nk} &= H(\rho_{AB}^{nk}\|\rho_A^{nk}\otimes\rho_B^{nk}) = H(\rho_A^{nk}) + H(\rho_B^{nk}) - H(\rho_{AB}^{nk}),\\
\tilde{H}_{nk} &= H(\rho_{AB}^{nk}\|\mu_n^{-1}\mathbf{P}_n^A\rho_A\mathbf{P}_n^A\otimes\eta_k^{-1}\mathbf{P}_k^B\rho_B\mathbf{P}_k^B)\\
&= -H(\rho_{AB}^{nk}) - \mathrm{tr}[\rho_A^{nk}\log(\mu_n^{-1}\mathbf{P}_n^A\rho_A\mathbf{P}_n^A)]\\
&\quad - \mathrm{tr}[\rho_B^{nk}\log(\eta_k^{-1}\mathbf{P}_k^B\rho_B\mathbf{P}_k^B)],
\end{aligned}$$

where

$$\mu_n = \mathrm{tr}[\mathbf{P}_n^A\rho_A],\quad \eta_k = \mathrm{tr}[\mathbf{P}_k^B\rho_B].$$

Using again Theorem 2.3.10, we have

$$\lim_{n,k\to+\infty} \tilde{H}_{nk} = \lim_{n,k\to+\infty} H((\mathbf{P}_n^A \otimes \mathbf{P}_k^B)\rho_{AB}(\mathbf{P}_n^A \otimes \mathbf{P}_k^B)\|(\mathbf{P}_n^A\rho_A\mathbf{P}_n^A) \otimes (\mathbf{P}_k^B\rho_B\mathbf{P}_k^B))$$
$$= H(\rho_{AB}\|\rho_A \otimes \rho_B).$$

Then we consider the difference between H_{nk} and $\tilde{H}_{nk}$; after some calculation, we obtain that it tends to zero:

$$\lim_{n,k\to\infty} (H_{nk} - \tilde{H}_{nk}) = \lim_{n,k\to\infty} (H(\rho_A^{nk}\|\mu_n^{-1}\mathbf{P}_n^A\rho_A\mathbf{P}_n^A) + H(\rho_B^{nk}\|\eta_k^{-1}\mathbf{P}_k^B\rho_B\mathbf{P}_k^B)) = 0.$$

The last double limit is equal to zero, since it follows from Theorem 2.3.10 that

$$\begin{aligned} 0 &\le H(\lambda_{nk}\rho_A^{nk}\|\mathbf{P}_n^A\rho_A\mathbf{P}_n^A) \\ &= H(\mathbf{P}_n^A \operatorname{tr}_B[\mathbf{I}_A \otimes \mathbf{P}_k^B\rho_{AB}\mathbf{I}_A \otimes \mathbf{P}_k^B]\mathbf{P}_n^A\|\mathbf{P}_n^A\rho_A\mathbf{P}_n^A) \\ &\le H(\operatorname{tr}_B[\mathbf{I}_A \otimes \mathbf{P}_k^B\rho_{AB}\mathbf{I}_A \otimes \mathbf{P}_k^B]\|\rho_A), \end{aligned}$$

implies

$$\lim_{n,k\to+\infty} H(\operatorname{tr}_B[\mathbf{I}_A \otimes \mathbf{P}_k^B\rho_{AB}\mathbf{I}_A \otimes \mathbf{P}_k^B]\|\rho_A) = H(\rho_A\|\rho_A) = 0.$$

Thus, the theorem on the limit of the intermediate sequence implies

$$\lim_{n,k\to+\infty} H(\lambda_{nk}\rho_A^{nk}\|\mathbf{P}_n^A\rho_A\mathbf{P}_n^A) = \lim_{n,k\to+\infty} H(\rho_A^{nk}\|\mu_n^{-1}\mathbf{P}_n^A\rho_A\mathbf{P}_n^A) = 0.$$

Similarly, we obtain that the second summand of the difference $H_{nk} - \tilde{H}_{nk}$ that also tends to zero:

$$\lim_{n,k\to+\infty} H(\rho_{AB}^{nk}\|\eta_k^{-1}\mathbf{P}_k^B\rho_B\mathbf{P}_k^B) = 0.$$

Finally, we have

$$\lim_{n,k\to+\infty} H(\rho_{AB}^{nk}\|\rho_A^{nk} \otimes \rho_B^{nk}) = H(\rho_{AB}\|\rho_A \otimes \rho_B)$$

and

$$\lim_{n,k\to+\infty} H(A_{nk}|B_{nk}) = H(A|B).$$

This proves the lemma. □

In the finite-dimensional case, conditional entropy has the properties of monotonicity, concavity in ρ_{AB} and subadditivity. These properties can be generalized to the infinite-dimensional case and will be derived in the following proposition due originally to Kuznetsova [102].

Proposition 8.3.6. *Suppose $H(\rho_A) < +\infty$. The conditional entropy function has the following properties:*

1. (Monotonicity) *The inequality*

$$H(A|BC)_{\rho_{ABC}} \leq H(A|B)_{\rho_{ABC}}, \tag{8.26}$$

holds for any systems A, B, C and for all $\rho_{ABC} \in \mathcal{S}(\mathbb{H}_{ABC})$;

2. (Concavity in ρ_{AB}) *If $\rho_{AB} = \alpha\rho^1_{AB} + (1-\alpha)\rho^2_{AB}$, $\alpha \in [0,1]$, then*

$$H(A|B)_{\rho_{AB}} \geq \alpha H(A^1|B^1) + (1-\alpha)H(A^2|B^2); \tag{8.27}$$

3. (Subadditivity of conditional entropy) *The inequality*

$$H(AB|CD) \leq H(A|C) + H(B|D) \tag{8.28}$$

holds for any systems A, B, C, D such that $H(\rho_A) < \infty$, $H(\rho_B) < \infty$.

Proof. (1) To prove monotonicity, we rewrite inequality (8.26) in the following way:

$$H(\rho_A) - H(\rho_{ABC}\|\rho_A \otimes \rho_{BC}) \leq H(\rho_A) - H(\rho_{AB}\|\rho_A \otimes \rho_B),$$

which is equivalent to

$$H(\rho_{AB}\|\rho_A \otimes \rho_B) \leq H(\rho_{ABC}\|\rho_A \otimes \rho_{BC}),$$

and the last inequality holds since the relative entropy in monotone with respect to taking the partial trace.

(2). To prove concavity, we will use the approximation by finite rank states. Consider again an arbitrary increasing sequence of finite rank projectors $\mathbf{P}^A_n$, $\mathbf{P}^B_k$, strongly converging to the operators $\mathbf{I}_A$, $\mathbf{I}_B$, respectively. Consider the sequence of states

$$\rho^{nk}_{AB} = \lambda^{-1}_{nk}((\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho_{AB}(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)), \quad \lambda_{nk} = \operatorname{tr}[(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho_{AB}]$$

with the partial states ρ^{nk}_A and ρ^{nk}_B. Let $\rho_{AB} = \alpha\rho^1_{AB} + (1-\alpha)\rho^2_{AB}$, $\alpha \in [0,1]$. Then

$$\begin{aligned}
\rho^{nk}_{AB} &= \lambda^{-1}_{nk}((\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho_{AB}(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)) \\
&= \frac{\alpha((\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho^1_{AB}(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)) + (1-\alpha)((\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho^2_{AB}(\mathbf{P}^A_n \otimes \mathbf{P}^B_k))}{\alpha \operatorname{tr}[\mathbf{P}^A_n \otimes \mathbf{P}^B_k\rho^1_{AB}] + (1-\alpha)\operatorname{tr}[\mathbf{P}^A_n \otimes \mathbf{P}^B_k\rho^2_{AB}]} \\
&= \frac{\theta^1_{nk}\rho^{1nk}_{AB} + \theta^2_{nk}\rho^{2nk}_{AB}}{\theta^1_{nk} + \theta^2_{nk}},
\end{aligned}$$

where

$$\theta^1_{nk} = \alpha\,\mathrm{tr}\big[(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho^1_{AB}\big], \quad \rho^{1nk}_{AB} = \frac{\alpha(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho^1_{AB}(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)}{\theta^1_{nk}},$$
$$\theta^2_{nk} = (1-\alpha)\,\mathrm{tr}\big[(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho^2_{AB}\big], \quad \rho^{2nk}_{AB} = \frac{(1-\alpha)(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)\rho^2_{AB}(\mathbf{P}^A_n \otimes \mathbf{P}^B_k)}{\theta^2_{nk}}.$$

Due to concavity of the conditional entropy on the finite rank states, we can write that

$$H(A_{nk}|B_{nk}) \geq \frac{\theta^1_{nk}}{\theta^1_{nk} + \theta^2_{nk}} H(A^1_{nk}|B^1_{nk}) + \frac{\theta^2_{nk}}{\theta^1_{nk} + \theta^2_{nk}} H(A^2_{nk}|B^2_{nk}).$$

Taking the limit and using Lemma 8.3.5 for both sides of the inequality, we obtain the proof of concavity.

(3). A direct verification shows that in the finite-dimensional case

$$H(AB|CD) = H(A|CD) + H(B|CD) - \big(H(A|CD) - H(A|BCD)\big), \tag{8.29}$$

which implies

$$H(AB|CD) \leq H(A|CD) + H(B|CD), \tag{8.30}$$

since the value in brackets in (8.29) is nonnegative. Using an approximation by finite rank states, we obtain inequality (8.30), on account of which we can immediately get (8.28) by using monotonicity of the conditional entropy.

Let $\rho_{ABCD} = \rho \in \mathcal{S}(\mathbb{H}_{ABCD})$ be a state with $H(\rho_A) < \infty, H(\rho_B) < \infty$. Consider an arbitrary increasing sequence of finite rank projectors $(\mathbf{P}^{CD}_l)^{+\infty}_{l=1}$, which strongly converges to the operator $\mathbf{I}_{CD}$. Consider also the sequence of states

$$\rho^l = \tau_l^{-1}(\mathbf{I}_A \otimes \mathbf{I}_B \otimes \mathbf{P}^{CD}_l)\rho(\mathbf{I}_A \otimes \mathbf{I}_B \otimes \mathbf{P}^{CD}_l), \quad \tau_l = \mathrm{tr}\big[(\mathbf{I}_A \otimes \mathbf{I}_B \otimes \mathbf{P}^{CD}_l)\rho\big].$$

The relation (8.29) (which implies (8.30)) holds for ρ^l since all the summands in (8.29) are finite. We will show that

$$H(A^l|CD^l) \to H(A|CD). \tag{8.31}$$

Actually, using Definition 8.3.1, we have

$$H(A^l|CD^l) = H(\rho^l_A) - H(\rho^l) - H(\rho^l_{ACD}\|\rho^l_A \otimes \rho^l_{CD}).$$

It follows from Theorem 2.3.10 that

$$\lim_{l\to\infty} H(\rho^l_A) = \lim_{l\to\infty} H(\tau_l^{-1}\,\mathrm{tr}_{BCD}\big[(\mathbf{I}_A \otimes \mathbf{I}_B \otimes \mathbf{P}^{CD}_l)\rho\big]) = H(\rho_A).$$

Furthermore, consider the values

$$H_l = H(\rho^l_{ACD}\|\rho^l_A \otimes \rho^l_{CD}) = H(\rho^l_{ACD}) + H(\rho^l_A) + H(\rho^l_{CD})$$

and

$$\begin{aligned}\tilde{H}_l &= H(\rho^l_{ACD}\|\rho_A \otimes \rho^l_{CD})\\ &= H(\tau^{-1}\mathbf{I}_A \otimes \mathbf{P}^{CD}_l \rho_{ACD}\mathbf{I}_A \otimes \mathbf{P}^l_{CD}\|\rho_A \otimes \tau_l^{-1}\mathbf{P}^{CD}_l \rho_{CD}\mathbf{P}^{CD}_l)\\ &= -H(\rho^l_{ACD}) + H(\rho^l_{CD}) + \operatorname{tr}[\tau_l^{-1}\operatorname{tr}_{CD}[\mathbf{I}_A \otimes \mathbf{P}^{CD}_l \rho_{ACD}](-\log\rho_A)].\end{aligned}$$

Then Theorem 2.3.10 implies that

$$\lim_{l\to+\infty} \tilde{H}_l = H(\rho_{ACD}\|\rho_A \otimes \rho_{CD}).$$

On the other hand, after some calculation and using Theorem 2.3.10, we obtain

$$\lim_{l\to+\infty}(\tilde{H}_l - H_l) = \lim_{l\to+\infty} H(\tau_l^{-1}\operatorname{tr}_{CD}[\mathbf{I}_A \otimes \mathbf{P}^{CD}_l \rho_{ACD}]\|\rho_A) = 0.$$

This implies that

$$\lim_{l\to+\infty} H(\rho^l_{ACD}\|\rho^l_A \otimes \rho^l_{CD}) = H(\rho_{ACD}\|\rho_A \otimes \rho_{CD})$$

and

$$\lim_{l\to+\infty} H(A^l|CD^l) = H(A|CD).$$

In a similar way, we obtain

$$\lim_{l\to+\infty} H(B^l|CD^l) = H(B|CD) \text{ and } \lim_{l\to+\infty} H(AB^l|CD^l) = H(AB|CD),$$

which implies that inequality (8.30) holds in an infinite-dimensional case, and hence, the subadditivity property of the conditional entropy holds. This proves the proposition. □

As mentioned in Lemma 7.2.10, the strong subadditivity of von Neumann entropy $H(\cdot)$ can be proved using the properties of conditional entropy.

Corollary 8.3.7. *Let* $\mathbb{H}_{123} := \mathbb{H}_1 \otimes \mathbb{H}_2 \otimes \mathbb{H}_3$ *and let* $\rho_{123} \in \mathcal{S}(\mathbb{H}_{123})$. *For* $i, j, k = 1, 2, 3$, *denote the reduced states* $\operatorname{tr}_{ij}[\rho_{123}] := \operatorname{tr}_{\mathbb{H}_i \otimes \mathbb{H}_j}[\rho_{123}]$ *by* $\rho_k = \operatorname{tr}_{ij}[\rho_{123}]$ *and* $\operatorname{tr}_k[\rho_{123}] := \operatorname{tr}_{\mathbb{H}_k}[\rho_{123}]$ *by* ρ_{ij}. *Then*

$$H(\rho_{123}) + H(\rho_2) \leq H(\rho_{12}) + H(\rho_{23}) \tag{8.32}$$

and

$$H(\rho_1) + H(\rho_2) \leq H(\rho_{13}) + H(\rho_{23}). \tag{8.33}$$

Proof. From Proposition 8.3.6, the conditional entropy $H(\rho_2|\rho_1)_\rho$ is concave. This implies the following inequality holds for $\rho_{123} \in \mathcal{S}(\mathbb{H}_{123})$:

$$H(\rho_1) + H(\rho_2) \le H(\rho_{13}) + H(\rho_{23}), \tag{8.34}$$

where $H(\rho_1) = \operatorname{tr}[\eta(\rho_1)] = -\operatorname{tr}[\rho_1 \log \rho_1]$, etc. This statement, somewhat similar to monotonicity, is obtained by considering

$$\Delta = \big(H(\rho_{13}) - H(\rho_1)\big) + \big(H(\rho_{23}) - H(\rho_2)\big).$$

The mapping $\rho_{123} \mapsto \rho_{13} := \operatorname{tr}_2[\rho_{123}]$ being linear, $H(\rho_{13}) - H(\rho_1)$ is concave in ρ_{123}. Similarly, the same is true for $H(\rho_{23}) - H(\rho_2)$. Therefore, Δ is concave in ρ_{123}. For pure states, $\Delta = 0$, since $H(\rho_{13}) = H(\rho_2)$ and $H(\rho_{23}) = H(\rho_1)$. By concavity, $\Delta \ge 0$ for mixed states. We now proceed by choosing a fourth Hilbert space $\mathbb{H}_4$ such that there is a pure state ρ_{1234} in $\mathbb{H}_{123} \otimes \mathbb{H}_4$, where $\rho_{123} = \operatorname{tr}_4[\rho_{1234}]$. Then

$$H(\rho_{123}) + H(\rho_2) - H(\rho_{12}) - H(\rho_{23}) = H(\rho_4) + H(\rho_2) - H(\rho_{12}) - H(\rho_{14}) \le 0$$

by (8.34). This proves the strong subadditivity. □

Obviously, it is possible to extend the domain of $H(A|B)$ to states ρ_{AB} such that $H(B) < +\infty$ or $H(AB) = H(C) < +\infty$. (Similarly, for the definition of conditional entropy, we use notation C for the reference system.) In this case, conditional entropy can be defined as in the finite-dimensional case (8.23), taking values $[-H(B), +\infty]$ provided $H(B) < +\infty$ or $[-\infty, H(C)]$ provided $H(C) < +\infty$. The above properties of conditional entropy (monotonicity, concavity and subadditivity) hold in the infinite-dimensional case. Monotonicity and subadditivity can be proved similarly for the finite-dimensional case by using the subadditivity of von Neumann entropy. To prove concavity, we use the following argument.

Consider density operator ρ_{AB} and spectral representations of its partial states

$$\rho_A = \sum_{i=1}^{+\infty} \alpha_i |i\rangle_A\langle i|, \quad \rho_B = \sum_{i=1}^{+\infty} \beta_j |j\rangle_B\langle j|,$$

where $\{|i\rangle_A, i = 1, 2, \ldots\}$, $\{|j\rangle,_B j = 1, 2, \ldots\}$ are the sets of eigenvectors of the states ρ_A and ρ_B in the spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. Also consider again increasing sequences of the spectral projectors $\mathbf{P}_n^A = \sum_{i=1}^n |i\rangle_A\langle i|$ and $\mathbf{P}_n^B = \sum_{j=1}^n |j\rangle_B\langle j|$. We also consider the state

$$\rho_{AB}^{nk} = \lambda_{nk}^{-1}(\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB} \mathbf{P}_n^A \otimes \mathbf{P}_k^B), \quad \lambda_{nk} = \operatorname{tr}[(\mathbf{P}_n^A \otimes \mathbf{P}_k^B)\rho_{AB}]$$

with its partial states ρ_A^{nk} and ρ_B^{nk}.

Then we prove concavity in a similar way to Proposition 8.3.6. Let $\rho_{AB} = \alpha\rho_{AB}^1 + (1-\alpha)\rho_{AB}^2$, $\alpha \in [0,1]$; then

$$\begin{aligned}\rho_{AB}^{nk} &= \lambda_{nk}^{-1}(\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB} \mathbf{P}_n^A \otimes \mathbf{P}_k^B)\\ &= \frac{\alpha(\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB}^1 \mathbf{P}_n^A \otimes \mathbf{P}_k^B) + (1-\alpha)(\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB}^2 \mathbf{P}_n^A \otimes \mathbf{P}_k^B)}{\alpha \operatorname{tr}[\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB}^1] + (1-\alpha)\operatorname{tr}[\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB}^2]}\\ &= \frac{\theta_{nk}^1 \rho_{AB}^{1nk} + \theta_{nk}^2 \rho_{AB}^{2nk}}{\theta_{nk}^1 + \theta_{nk}^2},\end{aligned}$$

where

$$\begin{aligned}\theta_{nk}^1 &= \alpha \operatorname{tr}[\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB}^1], \quad \rho_{AB}^{1nk} = \frac{\alpha(\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB}^1 \mathbf{P}_n^A \otimes \mathbf{P}_k^B)}{\theta_{nk}^1},\\ \theta_{nk}^2 &= \alpha \operatorname{tr}[\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB}^2], \quad \rho_{AB}^{2nk} = \frac{(1-\alpha)(\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB}^2 \mathbf{P}_n^A \otimes \mathbf{P}_k^B)}{\theta_{nk}^2},\end{aligned}$$

By using the concavity of the conditional entropy on finite rank states, we have

$$\begin{aligned}&H(\rho_{AB}^{nk}) - H(\rho_B^{nk})\\ &\geq \frac{\theta_{nk}^1}{\theta_{nk}^1 + \theta_{nk}^2}(H(\rho_{AB}^{1nk}) - H(\rho_B^{1nk})) + \frac{\theta_{nk}^2}{\theta_{nk}^1 + \theta_{nk}^2}(H(\rho_{AB}^{2nk}) - H(\rho_B^{2nk})).\end{aligned} \tag{8.35}$$

Note that $\lambda_{nk}\rho_B^{nk} = \mathbf{P}_k^B \operatorname{tr}_A[(\mathbf{P}_n^A \otimes \mathbf{I})\rho_{AB}(\mathbf{P}_n^A \otimes \mathbf{I})]\mathbf{P}_k^B \leq \mathbf{P}_k^B \rho_B \mathbf{P}_k^B$ for all k. For any $\varphi \in \mathbb{H}_B$, we have

$$\begin{aligned}&\sum_{i=1}^{+\infty} \langle i|_A \otimes \langle \varphi|_B (\mathbf{P}_n^A \otimes \mathbf{P}_k^B \rho_{AB} \mathbf{P}_n^A \otimes \mathbf{P}_k^B)|i\rangle_A \otimes |\varphi\rangle_B\\ &= \sum_{i=1}^{+\infty} \langle i|_A \otimes \langle \varphi|_B \left(\sum_{m=1}^{n} |m\rangle\langle m| \otimes \mathbf{P}_k^B \rho_{AB} \sum_{m'=1}^{n} |m'\rangle\langle m' \otimes \mathbf{P}_k^B \right) |i\rangle \otimes |\varphi\rangle\\ &= \sum_{i=1}^{n} \langle i| \otimes \langle \varphi \mathbf{P}_k^B|\rho_{AB}|i\rangle \otimes |\mathbf{P}_k^B \varphi\rangle\\ &\leq \sum_{i=1}^{+\infty} \langle i| \otimes \langle \varphi \mathbf{P}_k^B|\rho_{AB}|i\rangle \otimes |\mathbf{P}_k^B \varphi\rangle = \langle \varphi|\mathbf{P}_k^B \rho_B \mathbf{P}_k^B|\varphi\rangle.\end{aligned}$$

Also note that $H(\mathbf{P}_k^B \rho_B \mathbf{P}_k^B) < \infty$. Further, consider the inequality (8.35) and take the limit first on n and then on k. By applying Lemma 8.2.5 for the summands $H(\rho_{AB}^{nk})$, $H(\rho_{AB}^{i,nk})$, $i = 1, 2$ and (using the fact that $H(\mathbf{P}_k^B \rho_B \mathbf{P}_k^B) < +\infty$), and Lemma 8.2.5 for the summands $H(\rho_B^{\text{ink}})$, $H(\rho_B^{i,nk})$, $i = 1, 2$, we prove the concavity property.

9 Channel output entropies

In the following, let $\mathbb{H}_A$, $\mathbb{H}_B$ and $\mathbb{H}_E$ be separable complex Hilbert spaces that represent correspondingly input system A, output system B of a quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and environment system E that is interacting with A or B.

The purpose of this chapter is to explore various properties of the output entropy $H_\Phi(\cdot) := H(\Phi(\cdot)) : \mathcal{S}(\mathbb{H}_A) \to [0, +\infty]$ of the map $\Phi(\mathcal{S}(\mathbb{H}_A)) \subseteq \mathcal{S}(\mathbb{H}_B)$, where $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is a quantum channel. The properties of the output von Neumann entropy $H_\Phi(\cdot)$ of a quantum channel Φ reveal an important characteristic of this channel used in the study of its information properties.

9.1 Monotonicity of output relative entropy and Petz's theorem

The main results in this section consist of the monotonicity of channel output relative entropy (Theorem 9.1.3) that states

$$H\big(\Phi(\rho)\|\Phi(\sigma)\big) \le H(\rho\|\sigma), \quad \forall \Phi \in \mathfrak{QC}(A,B) \text{ and } \forall \rho,\sigma \in \mathcal{S}(\mathbb{H}_A),$$

(where $\mathfrak{QC}(A,B)$ denotes the class of quantum channels from A to B) and Petz's theorem (Theorem 9.1.5) that specifies the conditions under which the equality of the above inequality hold.

9.1.1 Monotonicity of output relative entropy

The proof of this result (Theorem 9.1.3), due originally to Lindblad [107], will be provided after the following lemmas.

Lemma 9.1.1. *Assume that $\mathbb{H}$ is a finite-dimensional complex Hilbert space. Assume that the quantum channel Φ takes the form of $\Phi(\mathbf{A}) = \sum_{i=1}^n \mathbf{V}_i\mathbf{A}\mathbf{V}_i^*$ for all $\mathbf{A} \in \mathfrak{B}(\mathbb{H})$, where $\mathbf{V}_i \in \mathfrak{B}(\mathbb{H})$ with $\sum_{i=1}^n \mathbf{V}_i\mathbf{V}_i^* = \mathbf{I}$ (i. e., Φ is completely positive and trace preserving). Then*

$$H\big(\Phi(\mathbf{A})\|\Phi(\mathbf{B})\big) \le H(\mathbf{A}\|\mathbf{B}), \quad \forall \mathbf{A},\mathbf{B} \in \mathfrak{T}_+(\mathbb{H}).$$

Proof. Assume that $\dim(\mathbb{H}) = n$. Put $\mathbb{K} = \mathbb{H} \otimes \mathbb{H}_n$, where $\mathbb{H}_n$ is any Hilbert space with $\dim(\mathbb{H}_n) = n$. Let $\{|i\rangle\}_{i=1}^n$ be a complete orthonormal set of $\mathbb{H}$ and $|\alpha\rangle$ be an arbitrary unit vector in $\mathbb{H}_n$. Define

$$\mathbf{W} = \sum_{i=1}^n \mathbf{V}_i \otimes \big(|i\rangle_{\mathbb{H}}\langle\alpha|_{\mathbb{H}_n}\big),$$

where $|i\rangle_{\mathbb{H}}\langle\alpha|_{\mathbb{H}_n} : \mathbb{H}_n \to \mathbb{H}$ is a map defined by

https://doi.org/10.1515/9783110788105-009

$$|i\rangle_{\mathbb{H}}\langle\alpha|_{\mathbb{H}_n}|\psi = \langle\alpha|\psi\rangle_{\mathbb{H}_n}|i\rangle_{\mathbb{H}}, \quad \forall\psi \in \mathbb{H}_n.$$

Then

$$\begin{aligned}\mathbf{W}^*\mathbf{W} &= \left(\sum_{i=1}^n \mathbf{V}_i \otimes (|i\rangle_{\mathbb{H}}\langle\alpha|_{\mathbb{H}_n})\right)^* \left(\sum_{i=1}^n \mathbf{V}_i \otimes (|i\rangle_{\mathbb{H}}\langle\alpha|_{\mathbb{H}_n})\right) \\ &= \sum_{i=1}^n (\mathbf{V}_i^*\mathbf{V}_i)\langle i|i\rangle_{\mathbb{H}} \otimes (|\alpha\rangle_{\mathbb{H}_n}\langle\alpha|) = \mathbf{I} \otimes (|\alpha\rangle_{\mathbb{H}_n}\langle\alpha|) = \mathbf{I}_{\mathbb{H}} \otimes \mathbf{P}_\alpha,\end{aligned}$$

where $\mathbf{P}_\alpha = |\alpha\rangle_{\mathbb{H}_n}\langle\alpha|$ is a one-dimensional projection on $\mathbb{H}_n$ along $|\alpha\rangle_{\mathbb{H}_n}$. Hence, there is a unitary operator $\mathbf{U}$ in $\mathbb{K}$ such that $\mathbf{W} = \mathbf{U} \otimes \mathbf{P}_\alpha$. Consequently,

$$\mathbf{U}(\mathbf{A} \otimes \mathbf{P}_\alpha)\mathbf{U}^* = \sum_{i,j=1}^n \mathbf{V}_i\mathbf{A}\mathbf{V}_j \otimes |i\rangle_{\mathbb{H}}\langle j|$$

and

$$\mathrm{tr}_2[\mathbf{U}(\mathbf{A} \otimes \mathbf{P}_\alpha)\mathbf{U}^*] = \sum_i \mathbf{V}_i\mathbf{A}\mathbf{V}_i^* = \Phi(\mathbf{A}),$$

where $\mathrm{tr}_2[\cdots]$ is the partial trace of $[\cdots]$ taken with respect to the second Hilbert space $\mathbb{H}_n$ of the tensor product. By Lemma 8.2.7, it follows that

$$\begin{aligned}H(\Phi(\mathbf{A})\|\Phi(\mathbf{B})) &= H(\mathrm{tr}_2[\mathbf{U}(\mathbf{A} \otimes \mathbf{P}_\alpha)\mathbf{U}^*] \| \mathrm{tr}_2[\mathbf{U}(\mathbf{B} \otimes \mathbf{P}_\alpha)\mathbf{U}^*]) \\ &\le H(\mathbf{U}(\mathbf{A} \otimes \mathbf{P}_\alpha)\mathbf{U}^* \| \mathbf{U}(\mathbf{B} \otimes \mathbf{P}_\alpha)\mathbf{U}^*) \quad \text{(by Lemma 8.2.7)} \\ &= H(\mathbf{A} \otimes \mathbf{P}_\alpha \| \mathbf{B} \otimes \mathbf{P}_\alpha) \quad \text{(since } \mathbf{U} \text{ is unitary)} \\ &\le H(\mathbf{A}\|\mathbf{B}) \quad \text{(since } \mathbf{P}_\alpha \text{ is a one-dimensional projection).}\end{aligned}$$

This proves the lemma. □

Lemma 9.1.2. *Let* $\Phi \in \mathfrak{QC}(A, B)$*, where* $\dim(\mathbb{H}_A) < +\infty$ *and* $\dim(\mathbb{H}_B) < +\infty$*. If* $\rho, \sigma \in \mathcal{S}(\mathbb{H}_A)$*, then*

$$H(\Phi(\rho)\|\Phi(\sigma)) \le H(\rho\|\sigma). \tag{9.1}$$

Proof. Kraus' representation theorem (4.4.4) implies that Φ can be approximated by finite linear combination

$$\Phi_n(\rho) = \sum_{i=1}^n \mathbf{V}_i\rho\mathbf{V}_i^*$$

with $\Phi_n^*(\mathbf{I}_B) = \sum_{i=1}^n \mathbf{V}_i^*\mathbf{V}_i \le \mathbf{I}_A$. Then

$$\Psi_n(\rho) \equiv \Phi_n(\rho) + \mathbf{V}_{n+1}\rho\mathbf{V}_{n+1}^*,$$

where $\mathbf{V}_{n+1} = \sqrt{\mathbf{I}_A - \Phi_n^*(\mathbf{I}_B)}$, is trace preserving. From Lemma 9.1.1, it follows that for all states ρ and σ,

$$H(\Psi_n(\rho)\|\Psi_n(\sigma)) \leq H(\rho\|\sigma).$$

Due to the fact that $\dim(\mathbb{H}_A)$ is finite, we obtain uniform convergence $\Psi_n(\rho)-\Phi(\rho) \to \mathbf{0}$ for every $\rho \in \mathcal{S}(\mathbb{H}_A)$. Obviously,

$$H(\Phi(\rho)\|\Phi(\sigma)) = \lim_{n\to+\infty} H(\Psi_n(\rho)\|\Psi_n(\sigma)) \leq H(\rho\|\sigma).$$

This proves the lemma. □

The following result establishes monotonicity of the relative entropy under a quantum channel $\Phi \in \mathfrak{QC}(A,B)$.

Theorem 9.1.3 (Lindblad [107]). *Let $\mathbb{H}_A$ and $\mathbb{H}_B$ be two separable complex Hilbert spaces, and let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel from A to B. Then*

$$H(\Phi(\rho)\|\Phi(\sigma)) \leq H(\rho\|\sigma), \quad \forall \rho,\sigma \in \mathcal{S}(\mathbb{H}_A). \tag{9.2}$$

Proof. Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be an increasing sequence of finite rank projections that converges strongly to the identity operator $\mathbf{I}_B$ on $\mathbb{H}_B$. Let $(\Phi_n)_{n=1}^{+\infty}$ be a sequence of a completely positive trace-preserving operator from $\mathfrak{T}_+(\mathbb{H}_A)$ to $\mathfrak{T}_+(\mathbb{H}_B)$ as in Lemma 9.1.2. From Lemmas 9.1.2 and 8.2.7, we know that for every n,

$$H(\Phi_n(\mathbf{A})\|\Phi_n(\mathbf{B})) \leq H(\mathbf{A}_n\|\mathbf{B}_n) \leq H(\mathbf{A}\|\mathbf{B}),$$

where $\mathbf{A}_n = \mathbf{P}_n\mathbf{A}$ and $\mathbf{B}_n = \mathbf{P}_n\mathbf{B}$ for $n = 1,2,\ldots$. From the lower semi-continuity of $H(\cdot\|\cdot)$ (see part 4 of Theorem 8.1.6), it follows that

$$H(\Phi(\mathbf{A})\|\Phi(\mathbf{B})) \leq \liminf_{n\to+\infty} H(\Phi_n(\mathbf{A})\|\Phi_n(\mathbf{B})) \leq H(\mathbf{A}\|\mathbf{B})$$

and the theorem is proved. □

9.1.2 Petz's theorem on output relative entropy

The conditions under which the monotonicity of relative entropy expressed in (9.2) become an equality that have been studied by many researchers including Hayden et al. [63], Jencova and Ruska [93], Petz [126], Ruskal [135] and Zhang and Wu [183].

We need the following lemma for our proof of Petz's theorem (Theorem 9.1.5).

Lemma 9.1.4. *Let $(\rho_n)_{n=1}^{+\infty}$ be a sequence of states in $\mathcal{S}(\mathbb{H})$ (where $\mathbb{H}$ is any separable Hilbert space) converging to a state ρ_0 and $(\mathbf{A}_n)_{n=1}^{+\infty}$ be a sequence of operators in the unit ball of $\mathfrak{B}(\mathbb{H})$ converging to an operator $\mathbf{A}_0$ in the weak operator topology (i. e., $\lim_{n\to+\infty}\langle\phi,\mathbf{A}_n\phi\rangle_{\mathbb{H}} = \langle\phi,\mathbf{A}_0\phi\rangle_{\mathbb{H}}$ for all $\phi \in \mathbb{H}$). Then the sequence $(\sqrt{\rho_n}\mathbf{A}_n\sqrt{\rho_n})_{n=1}^{+\infty}$ converges to $\sqrt{\rho_0}\mathbf{A}_0\sqrt{\rho_0}$ in the Hilbert–Schmidt norm $\|\cdot\|_{HS}$ defined in* (1.23).

Proof. Since the sequence $(\rho_n)_{n=1}^{+\infty}$ converges to ρ_0 in $\|\cdot\|_1$-norm, $\{\rho_n\}_{n=0}^{+\infty} \subset \mathcal{S}(\mathbb{H})$ is a compact set. The compactness criterion for subsets of $\mathcal{S}(\mathbb{H})$ (see Proposition 3.2.2) implies that for an arbitrary $\epsilon > 0$ there exists a finite- rank projector $\mathbf{P}_\epsilon$ on $\mathbb{H}$ such that $\operatorname{tr}[(\mathbf{I}_{\mathbb{H}} - \mathbf{P}_\epsilon)\rho_n] < \epsilon$ for all $n \geq 0$. We have

$$\begin{aligned}\sqrt{\rho_n}\mathbf{A}_n\sqrt{\rho_n} &= \sqrt{\rho_n}\mathbf{P}_\epsilon\mathbf{A}_n\mathbf{P}_\epsilon\sqrt{\rho_n} + \sqrt{\rho_n}\mathbf{P}_\epsilon\mathbf{A}_n(\mathbf{I}_{\mathbb{H}} - \mathbf{P}_\epsilon)\sqrt{\rho_n} \\ &\quad + \sqrt{\rho_n}(\mathbf{I}_{\mathbb{H}} - \mathbf{P}_\epsilon)\mathbf{A}_n\mathbf{P}_\epsilon\sqrt{\rho_n} + \sqrt{\rho_n}(\mathbf{I}_{\mathbb{H}} - \mathbf{P}_\epsilon)\mathbf{A}_n(\mathbf{I}_{\mathbb{H}} - \mathbf{P}_\epsilon)\sqrt{\rho_n}.\end{aligned} \tag{9.3}$$

Since $\mathbf{P}_\epsilon$ is a finite-rank operator on $\mathbb{H}$, $\mathbf{P}_\epsilon\mathbf{A}_n\mathbf{P}_\epsilon$ tends to $\mathbf{P}_\epsilon\mathbf{A}_0\mathbf{P}_\epsilon$ under the operator norm $\|\cdot\|_\infty$, and hence $\sqrt{\rho_n}\mathbf{P}_\epsilon\mathbf{A}_n\mathbf{P}_\epsilon\sqrt{\rho_n}$ tends to $\sqrt{\rho_0}\mathbf{P}_\epsilon\mathbf{A}_0\mathbf{P}_\epsilon\sqrt{\rho_0}$ in trace-norm $\|\cdot\|_1$. While it is easy to show that the Hilbert–Schmidt norm $\|\cdot\|_{HS}$ of the other terms of the right-hand side of (9.3) converges to zero as $\epsilon \to 0$ uniformly in n. This proves the lemma. □

The necessary and sufficient condition that guarantees the equality in (9.2) is stated below. Theorem 9.1.5 is originally proved in Petz [125] in the von Neumann algebra settings and with the transition probability instead of the relative entropy under the condition that ρ is full rank state in $\mathcal{S}(\mathbb{H}_A)$.

Theorem 9.1.5 (Petz's theorem [125]). *Let $\rho,\sigma \in \mathcal{S}(\mathbb{H}_A)$ be such that the relative entropy $H(\rho\|\sigma) < +\infty$. The equality holds in* (9.2) *if and only if $\Theta_\sigma(\Phi(\rho)) = \rho$, where Θ_σ is a channel from $\mathcal{S}(\mathbb{H}_B)$ to $\mathcal{S}(\mathbb{H}_A)$ defined by the formula*

$$\Theta_\sigma(\omega) = \sqrt{\sigma}\Phi^*\big((\Phi(\sigma))^{-1/2}(\omega)(\Phi(\sigma))^{-1/2}\big)\sqrt{\sigma}, \quad \forall\omega \in \mathcal{S}(\mathbb{H}_B). \tag{9.4}$$

Note that (9.4) implies that $\Theta_\sigma(\Phi(\sigma)) = \sigma$, so the above criterion for the equality in (9.2) can be treated as a reversibility condition (sufficiency) of the channel Φ with respect to the states ρ and σ in terms of Definition 5.6.1.

Proof of Theorem 9.1.5. A proof of the theorem for the finite-dimensional case can be found in Hial et al. [66], which will not be repeated here. For infinite-dimensional $\mathbb{H}_A$ and $\mathbb{H}_B$, we follow the proof provided in the Appendix of Shirokov [149] below.

(⇒) We assume that $\Theta_\sigma(\Phi(\rho)) = \rho$, i. e., the channel Φ is reversible with respect to the states ρ and σ. The equality $H(\Phi(\rho)\|\Phi(\sigma)) = H(\rho\|\sigma)$ has been established in Shirokov [149] and will not be repeated here.

(⇐) Assume that $H(\Phi(\rho)\|\Phi(\sigma)) = H(\rho\|\sigma)$ for all $\rho,\sigma \in \mathcal{S}(\mathbb{H}_A)$. We want to show that $\Theta_\sigma(\omega) = \sqrt{\sigma}\Phi^*((\Phi(\sigma))^{-1/2}(\omega)(\Phi(\sigma))^{-1/2})\sqrt{\sigma}$, $\forall\omega \in \mathcal{S}(\mathbb{H}_B)$. We first note that $H(\rho\|\sigma) < +\infty$ does not imply that $\lambda\rho \leq \sigma$ for some $\lambda > 0$, and hence the argument

$(\cdots)$ of the map $\Phi^*(\cdots)$ in (9.4) with $\omega = \Phi(\rho)$ may be an unbounded operator. Nevertheless, we can define the channel Θ_σ as a predual map to the linear completely positive unital map

$$\Theta_\sigma(\zeta) = [\Phi(\sigma)]^{-1/2}\Phi([\sigma]^{1/2}(\zeta)[\sigma]^{1/2})[\Phi(\sigma)]^{-1/2}, \quad \forall\zeta \in \mathcal{S}(\mathbb{H}_A). \tag{9.5}$$

This means that we can use formula (9.4), keeping in mind that Φ^* is an extension of the dual map to unbounded operators in $\mathbb{H}_B$ (which can be defined by $\Phi^*(\cdot) = \sum_k \mathbf{V}_k^*(\cdot)\mathbf{V}_k$ via the Kraus representation $\Phi(\cdot) = \sum_k \mathbf{V}_k(\cdot)\mathbf{V}_k^*$). With this definition of the channel Θ_σ with the nonfull rank state ρ, we will prove the following. Consider the ensemble consisting of two states ρ and σ with probabilities t and $1-t$, where $t \in]0,1[$. Let $\sigma_t = t\rho + (1-t)\sigma$. By Donald''s identity (part 7 of Theorem 8.1.6), we have

$$tH(\rho\|\sigma) + (1-t)H(\sigma\|\sigma) = tH(\rho\|\sigma_t) + (1-t)H(\sigma\|\sigma_t) + H(\sigma_t\|\sigma) \tag{9.6}$$

and

$$\begin{aligned} &tH(\Phi(\rho)\|\Phi(\sigma)) + (1-t)H(\Phi(\sigma)\|\Phi(\sigma)) \\ &\quad = tH(\Phi(\rho)\|\Phi(\sigma_t)) + (1-t)H(\Phi(\sigma)\|\Phi(\sigma_t)) + H(\Phi(\sigma_t)\|\Phi(\sigma)) \\ &\quad \le tH(\rho\|\sigma_t) + (1-t)H(\sigma\|\sigma_t) + H(\sigma_t\|\sigma) \\ &\quad = tH(\rho\|\sigma) + (1-t)H(\sigma\|\sigma), \end{aligned} \tag{9.7}$$

where the left-hand sides are finite and coincide by the condition. Since the first, the second and the third terms in the right-hand side of (9.7) are not less than the corresponding terms in (9.8) by monotonicity of the relative entropy, we obtain

$$H(\Phi(\rho)\|\Phi(\sigma_t)) = H(\rho\|\sigma_t) \quad \text{and} \quad H(\Phi(\sigma)\|\Phi(\sigma_t)) = H(\sigma\|\sigma_t). \tag{9.8}$$

Consequently, we have $\rho = \Theta_t(\Phi(\rho))$ for all $t \in]0,1[$, where

$$\Theta_t(\rho) = [\sigma_t]^{1/2}\Phi^*([\Phi(\sigma_t)]^{-1/2}(\rho)[\Phi(\sigma_t)]^{-1/2})[\sigma_t]^{1/2}, \quad \forall\rho \in \mathcal{S}(\mathbb{H}_B).$$

To complete the proof, it suffices to show that

$$\lim_{t\to 0}\Theta_t = \Theta_\sigma \tag{9.9}$$

in the strong convergence topology (in which $\Phi_n \to \Phi$ strongly means $\Phi_n(\rho) \to \Phi(\rho)$ for all ρ), since this implies $\rho = \lim_{t\to+0}\Theta_t(\Phi(\rho)) = \Theta_\sigma(\Phi(\rho))$. Since $\Theta_t(\Phi(\sigma)) = \sigma$ for all $t \in]0,1[$, the set of channels $\{\Theta_t\}_{t\in]0,1[}$ is relatively compact in the strong convergence topology. Hence, there exists a sequence $(t_n)_{n=1}^{+\infty}$ converging to zero such that

$$\lim_{n\to+\infty}\Theta_{t_n} = \Theta_0, \tag{9.10}$$

where Θ_0 is a particular channel. We will show that $\Theta_0 = \Theta_\sigma$. Note that (9.10) means that the sequence $(\Theta^*_{t_n}(\mathbf{A})$ (where $\Theta^*_{t_n}(\cdot) : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ is the dual channel of $\Theta_{t_n}(\cdot)$) tends to the operator $\Theta^*_0(\mathbf{A})$ in the weak operator topology for any positive $\mathbf{A} \in \mathfrak{B}(\mathbb{H}_B)$. By Lemma (9.1.4), we have

$$\lim_{n\to+\infty}[\Phi(\sigma_{t_n})]^{1/2}\Theta^*_{t_n}(\mathbf{A})[\Phi(\sigma_{t_n})]^{1/2} = [\Phi(\sigma)]^{1/2}\Theta^*_0(\mathbf{A})[\Phi(\sigma)]^{1/2}$$

in the Hilbert–Schmidt norm $\|\cdot\|_{HS}$. But the explicit form of $\Theta^*_{t_n}$ shows that

$$[\Phi(\sigma_{t_n})]^{1/2}\Theta^*_{t_n}(\mathbf{A})[\Phi(\sigma_{t_n})]^{1/2} = \Phi([\sigma]^{1/2}\mathbf{A})[\sigma]^{1/2})$$

and since $\lim_{n\to+\infty}[\sigma_{t_n}]^{1/2}\mathbf{A}[\sigma_{t_n}]^{1/2} = [\sigma]^{1/2}\mathbf{A}[\sigma]^{1/2}$ in the trace-norm $\|\cdot\|_1$, the above limit coincides with $\Phi([\sigma]^{1/2}\mathbf{A}[\sigma]^{1/2})$. So, we have $\Theta^*_0(\mathbf{A}) = \Theta^*_\sigma(\mathbf{A})$ for all $\mathbf{A}$, and hence $\Theta_0 = \Theta_\sigma$. The above observation shows that for an arbitrary sequence $(t_n)_{n=1}^{+\infty}$ converging to zero any partial limit of the sequence $(\Theta_{t_n})_{n=1}^{+\infty}$ coincides with Θ_σ, which means (9.9). This proves Petz's theorem. □

9.2 Continuity of output entropies

In this section, we investigate continuity of the output entropy $H_\Phi(\cdot) := H(\Phi(\cdot)) : \mathfrak{T}_+(\mathbb{H}_A) \to [0, +\infty]$, when the map $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ is:
(i) a general positive linear map;
(ii) a quantum operation; or
(iii) a quantum channel.

These three cases are to be explored in the following subsections.

9.2.1 Positive linear map Φ

Let $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ be a positive linear map. Since the von Neumann entropy $H(\cdot) : \mathcal{S}(\mathbb{H}_A) \to [0, +\infty]$ is a concave and lower semicontinuous function, it is easy to see that the output von Neumann entropy $H_\Phi(\cdot) := (H \circ \Phi)(\cdot) = H(\Phi(\cdot))$ of the positive linear map Φ is also a concave nonnegative lower semicontinuous function on the set $\mathcal{S}(\mathbb{H}_A) \subset \mathfrak{T}_+(\mathbb{H}_A)$.

We need the following lemmas for proving the main result (Theorem 9.2.3) in this subsection.

Lemma 9.2.1. *Assume that $\mathbb{H}_A = \mathbb{H}_B = \mathbb{H}$. For an arbitrary convex set $\mathcal{A} \subset \mathcal{S}(\mathbb{H})$, on which the von Neumann entropy $H(\cdot)$ is bounded. Then there exists an operator $\mathbf{T} \in \mathfrak{B}(\mathbb{H})$ such that*

$$\sup_{\mathbf{A}\in\mathcal{A}} \operatorname{tr}\big[\mathbf{A}(-\log \mathbf{T})\big] < +\infty \quad \textit{and} \quad \mathbf{UT} = \mathbf{TU}$$

for any unitary $\mathbf{U} \in \mathfrak{B}(\mathbb{H})$ *such that* $\mathbf{UAU}^* \in \mathcal{A}$ *for all* $\mathbf{A} \in \mathcal{A}$.

Proof. Let

$$\mathbb{K} = \mathbb{C}|0\rangle_{\mathbb{H}} = \{c|0\rangle_{\mathbb{H}} \mid c \in \mathbb{C}\}$$

be the one-dimensional subspace of $\mathbb{H}$ generated by the unit vector $|0\rangle_{\mathbb{H}}$. Consider the convex set $\mathcal{A}^e = \{\rho_{\mathbf{A}} \mid \mathbf{A} \in \mathcal{A}\}$ of states in $\mathcal{S}(\mathbb{H} \oplus \mathbb{K})$, where $\rho_{\mathbf{A}}$ is defined by

$$\rho_{\mathbf{A}} = \mathbf{A} \oplus (1 - \operatorname{tr}[\mathbf{A}])|0\rangle_{\mathbb{H}}\langle 0|.$$

For arbitrary $\mathbf{A} \in \mathcal{A}$, we have

$$\begin{aligned} H(\rho_{\mathbf{A}}) &= \operatorname{tr}\big[\eta(\rho_{\mathbf{A}})\big] \\ &= \operatorname{tr}\big[\eta(\mathbf{A} \oplus (1 - \operatorname{tr}[\mathbf{A}])|0\rangle_{\mathbb{H}}\langle 0|)\big] = \operatorname{tr}\big[\eta(\mathbf{A}) + \eta((1 - \operatorname{tr}[\mathbf{A}])|0\rangle_{\mathbb{H}}\langle 0|)\big] \\ &= \operatorname{tr}\big[\eta(\mathbf{A})\big] + \operatorname{tr}\big[\eta((1 - \operatorname{tr}[\mathbf{A}])(|0\rangle_{\mathbb{H}}\langle 0|))\big] \\ &= H(\mathbf{A}) + \eta\big(\operatorname{tr}[\mathbf{A}]\big) + \eta\big(1 - \operatorname{tr}[\mathbf{A}]\big) = H(\mathbf{A}) + 1, \end{aligned}$$

where $\eta : [0, +\infty] \to [-\infty, +\infty]$ is defined by

$$\eta(x) = \begin{cases} -x \log x, & \text{if } x > 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Thus, the von Neumann entropy $H(\cdot)$ is bounded on the convex set $\mathcal{A}^e$. Hence, the Holevo χ-capacity defined by

$$C_\chi(\mathcal{A}^e) = \sup_{\mu\in\mathcal{P}_{\mathcal{A}^e}} \chi(\mu) = \sup_{\mu\in\mathcal{P}_{\mathcal{A}^e}} \bigg(\int_{\mathcal{S}(\mathbb{H}_A)} H(\rho\|\bar{\rho}(\mu))\mu(d\rho) \bigg) < +\infty,$$

where

$$\mathcal{P}_{\mathcal{A}^e} := \mathcal{P}_{\mathcal{A}^e}\big(\mathcal{S}(\mathbb{H}_A)\big) = \big\{\mu \in \mathcal{P}\big(\mathcal{S}(\mathbb{H}_A)\big) \mid \bar{\rho}(\mu) \in \mathcal{A}^e\big\}$$

(see Chapter 12 for the definition of χ-capacity). This implies existence of the unique state $\omega(\mathcal{A}^e)$ in $\overline{\mathcal{A}^e}$ (called the optimal average state of the set $\mathcal{A}^e$) such that

$$H(\rho\|\omega(\mathcal{A}^e)) \le C_\chi(\mathcal{A}^e)$$

for all $\rho \in \mathcal{A}^e$. The state $\omega(\mathcal{A}^e)$ has the form $\mathbf{T} \oplus |0\rangle_{\mathbb{H}}\langle 0|$, where $\mathbf{T} \in \mathfrak{Q}\mathfrak{C}(\mathbb{H})$ and $\lambda > 0$. For arbitrary unitary $\mathbf{U}$ in $\mathfrak{B}(\mathbb{H})$, such that $\mathbf{U}\mathcal{A}\mathbf{U}^* = \mathcal{A}$. Consequently, $(\mathbf{U} \oplus \mathbf{I}_{\mathbb{K}})\omega(\mathcal{A}^e) = \omega(\mathcal{A}^e)(\mathbf{U} \oplus \mathbf{I}_{\mathbb{K}})$, and hence $\mathbf{UT} = \mathbf{TU}$. This proves the lemma. □

The following proposition is a summary of results in Proposition 7.3.7 that will be useful in establishing the results in output entropy. Recall from (7.12) that the increasing number $\lambda^\dagger(\mathbf{H})$ of an $\mathfrak{H}$-operator $\mathbf{H}$ is defined as $\lambda^\dagger(\mathbf{H}) = \inf\{\lambda > 0 \mid \mathrm{tr}[\exp(-\lambda\mathbf{H})] < +\infty\}$ and $\mathcal{K}_{\mathbf{H}}(h)$ is a compact convex subset of $\mathcal{S}(\mathbb{H}_A)$. We have used the convention that $\lambda^\dagger(\mathbf{H}) = +\infty$ if $\mathrm{tr}[\exp(-\lambda\mathbf{H})] = +\infty$ for any $\lambda > 0$ and $\lambda^\dagger(\mathbf{H}) = 0$ if $\mathrm{tr}[\exp(-\lambda\mathbf{H})] < +\infty$ for all $\lambda > 0$.

Proposition 9.2.2. *Let $\mathbf{H}$ be an $\mathfrak{H}$-operator on a separable complex Hilbert space $\mathbb{H}$ and $h > 0$.*
1. *The von Neumann entropy $H(\cdot)$ is bounded on the set $\mathcal{K}_{\mathbf{H}}(h)$ if and only if $\lambda^\dagger(\mathbf{H}) < +\infty$;*
2. *The von Neumann entropy $H(\cdot)$ is continuous on the set $\mathcal{K}_{\mathbf{H}}(h)$ if and only if $\lambda^\dagger(\mathbf{H}) = 0$.*

The following theorem (due originally to Shirokov [147]) shows that the output von Neumann entropy $H_\Phi(\cdot) := H(\Phi(\cdot))$ cannot be finite and discontinuous simultaneously for any positive (but not necessarily completely positive) linear map $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ that is trace nonincreasing.

The following theorem is due originally to Shirokov [147].

Theorem 9.2.3 (Shirokov [147]). *Let Φ be a positive linear map from $\mathfrak{T}_+(\mathbb{H}_A)$ to $\mathfrak{T}_+(\mathbb{H}_B)$ that is trace nonincreasing. The following properties are equivalent:*
1. *the function $\rho \mapsto H_\Phi(\rho)$ is finite on $\mathcal{S}(\mathbb{H}_A)$;*
2. *the function $\rho \mapsto H_\Phi(\rho)$ is continuous and bounded on $\mathcal{S}(\mathbb{H}_A)$;*
3. *there exists an orthonormal basis $\{|i\rangle_B\}_{i=1}^{+\infty}$ of the space $\mathbb{H}_B$ such that the function $\rho \mapsto H(\{\langle i|\Phi(\rho)|i\rangle_B\}_{i=1}^{+\infty})$ is continuous and bounded on $\mathcal{S}(\mathbb{H}_A)$;*
4. *there exists an orthonormal basis $\{|i\rangle_B\}_{i=1}^{+\infty}$ of the space $\mathbb{H}_B$ and a sequence $(h_i)_{i=1}^{+\infty}$ of nonnegative numbers such that*

$$\left\| \sum_{i=1}^{+\infty} h_i \Phi^*(|i\rangle_B\langle i|) \right\|_\infty < +\infty \quad \text{and} \quad \sum_{i=1}^{+\infty} \exp(-h_i) < +\infty,$$

where $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ is the dual map to Φ.

The set $\mathcal{S}(\mathbb{H}_A)$ in (1) can be replaced by arbitrary convex closed bounded subset $\mathcal{A} \subset \mathfrak{T}_+(\mathbb{H}_A)$ such that

$$\sup_{\mathbf{A}\in\mathcal{A}}\Big(\lim_{n\to+\infty} \mathrm{tr}[\mathbf{A}\mathbf{B}_n]\Big) < +\infty \Rightarrow \sup_n \|\mathbf{B}_n\|_\infty < +\infty$$

for any increasing sequence $(\mathbf{B}_n)_{n=1}^{+\infty}$ of positive operators in $\Phi^*(\mathfrak{B}(\mathbb{H}_B))$.

Proof. (1) $\Rightarrow$ (2) Validity of the discrete Jensen inequality for the concave finite nonnegative function $\rho \mapsto H_\Phi(\rho)$ implies its boundedness (see Proposition 3.4.4). Indeed, if the opposite were true that for each natural n there exists a state ρ_n such that $H_\Phi(\rho_n) \geq 2^n$ (for contradiction purpose), then by discrete Jensen inequality, we have

$$H_\Phi\left(\sum_{n=1}^{+\infty} 2^{-n}\rho_n\right) \geq \sum_{n=1}^{+\infty} 2^{-n} H_\Phi(\rho_n) = +\infty.$$

Lemma 9.2.1 implies existence of an $\mathfrak{H}$-operator $\mathbf{H} = -\log \mathbf{T}$ such that $\lambda^\dagger(\mathbf{H}) < +\infty$ and $\operatorname{tr}[\mathbf{H}\Phi(\rho)] \leq h$ for all $\rho \in \mathcal{S}(\mathbb{H}_A)$ and some $h > 0$. Let $\mathbf{H} = \sum_{i=1}^{+\infty} h_i |i\rangle_B\langle i|$. Since the function

$$\begin{aligned}\rho \mapsto \operatorname{tr}[\mathbf{H}\Phi(\rho)] &= \operatorname{tr}\left[\left(\sum_{i=1}^{+\infty} h_i |i\rangle_B\langle i|\right)\Phi(\rho)\right] \\ &= \sum_{i=1}^{+\infty} h_i \langle i|\Phi(\rho)|i\rangle_B = \operatorname{tr}\left[\sum_{i=1}^{+\infty} h_i \Phi^*(|i\rangle_B\langle i|)\rho\right]\end{aligned}$$

is bounded on $\mathcal{S}(\mathbb{H}_B)$, the linear operator $\sum_{i=1}^{+\infty} h_i\Phi^*(|i\rangle_B\langle i|)$ is bounded in $\mathbb{H}_A$ under the operator norm $\|\cdot\|_\infty$. Thus, the above function is continuous on $\mathcal{S}(\mathbb{H}_A)$. For arbitrary compact subset $\mathcal{K}$ of $\mathcal{S}(\mathbb{H}_A)$, Dini's lemma (see Rudin [134]) implies uniform convergence of the infinite series of operators $\sum_{i=1}^{+\infty} h_i \operatorname{tr}[\Phi^*(|i\rangle_B\langle i|)\rho]$ on the set $\mathcal{K} \subset \mathcal{S}(\mathbb{H}_A)$, and hence, existence of a nondecreasing sequence $(y_i^{\mathcal{K}})_{i=1}^{+\infty}$ of positive numbers converging to $+\infty$ such that

$$\sup_{\rho\in\mathcal{K}}\left(\sum_{i=1}^{+\infty} y_i^{\mathcal{K}} h_i \operatorname{tr}[\Phi^*(|i\rangle_B\langle i|)\rho]\right) < +\infty.$$

Let $\mathbf{H}^{\mathcal{K}} = \sum_{i=1}^{+\infty} h_i |i\rangle_B\langle i|$ be an $\mathfrak{H}$-operator on $\mathbb{H}_B$ with $\lambda^\dagger(\mathbf{H}^{\mathcal{K}}) = 0$. Thus, we have

$$\sup_{\rho\in\mathcal{K}}(\operatorname{tr}[\mathbf{H}^{\mathcal{K}}\Phi(\rho)]) = \sup_{\rho\in\mathcal{K}}\left(\sum_{i=1}^{+\infty} y_i^{\mathcal{K}} h_i \operatorname{tr}[\Phi^*(|i\rangle_B\langle i|)\rho]\right) < +\infty. \tag{9.11}$$

By part (2) of Proposition 7.3.7, the function $\rho \mapsto H(\Phi(\rho))$ is continuous on the set $\mathcal{K}$, and hence on the set $\mathcal{S}(\mathbb{H}_A)$ (since $\mathcal{K}$ is an arbitrary compact subset of $\mathcal{S}(\mathbb{H}_A)$).

(1) $\Rightarrow$ (4). In the proof of (1) $\Rightarrow$ (2), existence of the basis $\{|i\rangle_B\}_{i=1}^{+\infty}$ and of the sequence $(h_i')_{i=1}^{+\infty}$, where $h_i' = \lambda h_i$, $\lambda > 0$, with the desired properties is shown.

(4) $\Rightarrow$ (3) follows from the proof of (1) $\Rightarrow$ (2), since (9.11) and part (2) of Proposition 7.3.7 implies continuity of the function $\rho \mapsto H(\{\langle i|\Phi(\rho)|i\rangle_B\}_{i=1}^{+\infty})$ on the set $\mathcal{K}$.

(3) $\Rightarrow$ (1) follows from relation $H(\rho) \leq H(\{\langle i|\Phi(\rho)|i\rangle_B\}_{i=1}^{+\infty})$.

The last assertion of the theorem is a corollary of the proof of (1) $\Rightarrow$ (2).

This proves the theorem. □

Remark 9.1. Theorem 9.2.3 does not assert that finiteness of the von Neumann entropy $H(\cdot)$ on the set $\Phi(\mathcal{S}(\mathbb{H}_A)) \subseteq \mathcal{S}(\mathbb{H}_B)$ implies its continuity on this set, since continuity of the function $\rho \mapsto H(\Phi(\rho))$ on the noncompact set $\mathcal{S}(\mathbb{H}_A)$ does not imply continuity of the function $\mathbf{B} \mapsto H(\mathbf{B})$ on the set $\Phi(\mathcal{S}(\mathbb{H}_A)) \subseteq \mathcal{S}(\mathbb{H}_B)$. To show this, we consider the following example. Let $\mathcal{B}$ be a convex closed subset of $\mathcal{S}(\mathbb{H}_B)$ on which the von Neumann entropy $H(\cdot)$ is bounded but not continuous. Let $(\sigma_n)_{n=1}^{+\infty}$ be a sequence of states in $\mathcal{B}$ converging to a state σ_0 such that $\lim_{n\to+\infty} H(\sigma_n) \neq H(\sigma_0)$. Consider the map $\Phi : \rho \mapsto \sum_{n=0}^{+\infty} \langle n|\rho|n\rangle_A \sigma_n$, where $(|n\rangle_A)_{n=0}^{+\infty}$ is a particular orthonormal basis in $\mathbb{H}_A$. By Theorem 9.2.3, the function $\rho \mapsto H_\Phi(\rho)$ is continuous on the set $\mathcal{S}(\mathbb{H}_A)$ but the function $\mathbf{B} \mapsto H(\mathbf{B})$ is not continuous on the set $\Phi(\mathcal{S}(\mathbb{H}_A)) \subseteq \mathcal{S}(\mathbb{H}_B)$ containing the sequence $(\sigma_n)_{n=1}^{+\infty}$ and the state σ_0. Continuity of the function $\rho \mapsto H_\Phi(\rho)$ on the set $\mathcal{S}(\mathbb{H}_A)$ means continuity of the function $\mathbf{B} \mapsto H(\mathbf{B})$ on each set of the form $\Phi(\mathcal{K}) \subset \mathcal{S}(\mathbb{H}_B)$, where $\mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{H}_A)$.

Remark 9.2. The main assertion of Theorem 9.2.3 (the implication (1) ⇒ (2)) is based on the specific property of the von Neumann entropy; it cannot be proved by using only such general properties of entropy-type functions as concavity, lower semicontinuity and nonnegativity. The simplest example showing this is given by the function $\rho \mapsto R_0(\Phi(\rho)) = \|\Phi(\rho)\|_\infty \log(\mathrm{rank}(\Phi(\rho)))$, the output 0-order Renyi entropy of the map Φ. Theorem 9.2.3 can be used to obtain a condition of continuity of the output entropy for a certain class of quantum channels.

Theorem 9.2.3 and inequality (7.48) imply the following observation (which can be directly proved by using Lemma 7.4.3).

Corollary 9.2.4. *Let Φ and Ψ be positive linear maps from $\mathfrak{T}_+(\mathbb{H}_A)$ to $\mathfrak{T}_+(\mathbb{H}_B)$ that are trace nonincreasing, and let $\lambda \in]0,1[$. The map $\lambda\Phi + (1-\lambda)\Psi$ has continuous output entropy if and only if the maps Φ and Ψ have continuous output entropies $H(\Phi(\cdot)), H(\Psi(\cdot)) : \mathcal{S}(\mathbb{H}_A) \to [0,+\infty]$.*

Thus, the set of all positive maps with continuous output entropy is convex and forms a *face* of the convex set $\mathfrak{L}_{\leq 1}^{+}(\mathbb{H}_A, \mathbb{H}_B)$, where $\mathfrak{L}_{\leq 1}^{+}(\mathbb{H}_A, \mathbb{H}_B)$ is the set of all positive linear maps from $\mathfrak{T}_+(\mathbb{H}_A)$ to $\mathfrak{T}_+(\mathbb{H}_B)$ that are trace non-increasing. It is easy to show that this face is dense in $\mathfrak{L}_{\leq 1}^{+}(\mathbb{H}_A, \mathbb{H}_B)$ in the strong convergence topology.

We will use the following corollary of Theorem 9.2.3 and inequality (7.49).

Corollary 9.2.5. *Let $\{\Phi_i\}_{i\in I}$ be a finite or countable family of positive linear maps from $\mathfrak{T}_+(\mathbb{H}_A)$ to $\mathfrak{T}_+(\mathbb{H}_B)$ that are trace nonincreasing such that*

$$\sup_{\rho\in\mathcal{S}(\mathbb{H}_A)} \sum_{i\in I} \mathrm{tr}[\Phi_i(\rho)] < +\infty.$$

Then the output entropy of the map $\sum_{i\in I} \Phi_i$ is continuous if

$$\sum_{i\in I} H(\Phi_i(\rho)) < +\infty \quad \textit{and} \quad H(\{\mathrm{tr}[\Phi_i(\rho)]\}_{i\in I}) < +\infty, \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A).$$

This condition is necessary if either $\mathrm{supp}(\Phi_i(\rho)) \perp \mathrm{supp}(\Phi_j(\rho))$ *for all* $i \neq j$ *and for all* $\rho \in \mathcal{S}(\mathbb{H}_A)$ *or* $\mathbb{I}$ *is a finite set.*

Theorem 9.2.3 provides a simple proof for the following corollary.

Corollary 9.2.6. *Let* $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ *and* $\Psi : \mathfrak{T}_+(\mathbb{K}_A) \to \mathfrak{T}_+(\mathbb{K}_B)$ *be two positive bounded linear maps having continuous output entropies* $H_\Phi(\cdot)$ *and* $H_\Psi(\cdot)$*, respectively. If the map* $\Phi \otimes \Psi : \mathfrak{T}_+(\mathbb{H}_A \otimes \mathbb{K}_A) \to \mathfrak{T}_+(\mathbb{H}_B \otimes \mathbb{K}_B)$ *is positive, then it has continuous output entropy* $H_{\Phi\otimes\Psi}(\cdot) : \mathfrak{T}_+(\mathbb{H}_A \otimes \mathbb{K}_A) \to [0, +\infty]$.

Proof. Without loss of generality, we may assume that the maps Φ and Ψ are trace nonincreasing. By Theorem 9.2.3, it is sufficient to prove that $H((\Phi \otimes \Psi)(\omega)) < +\infty$ for any $\omega \in \mathfrak{T}_+(\mathbb{H}_A \otimes \mathbb{K}_A)$. This follows from subadditivity of the quantum entropy since $\mathrm{tr}_{\mathbb{K}_A}[(\Phi \otimes \Psi)(\omega)] \leq \Phi(\mathrm{tr}_{\mathbb{K}_A}[\omega])$ and $\mathrm{tr}_{\mathbb{H}_A}[(\Phi \otimes \Psi)(\omega)] \leq \Psi(\mathrm{tr}_{\mathbb{H}_A}[\omega])$. This proves the corollary. ☐

Corollary 9.2.7. *Let* $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ *and* $\Psi : \mathfrak{T}_+(\mathbb{K}_A) \to \mathfrak{T}_+(\mathbb{K}_B)$ *be positive linear bounded maps such that the map* $\Phi \otimes \Psi : \mathfrak{T}_+(\mathbb{H}_A \otimes \mathbb{K}_A) \to \mathfrak{T}_+(\mathbb{H}_B \otimes \mathbb{K}_B)$ *is positive. If the map* Ψ *is trace preserving and has finite (and hence continuous) output entropy* $H_\Psi(\cdot) := H(\Psi(\cdot))$*, then the following properties are equivalent:*
1. $H((\Phi \otimes \Psi)(|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|)) < +\infty$ *for any unit vector* $\phi \in \mathbb{H}_A \otimes \mathbb{K}_A$*;*
2. *the map* Φ *has continuous output entropy* $H_\Phi(\cdot)$*;*
3. *the map* $\Phi \otimes \Psi$ *has continuous output entropy* $H_{\Phi\otimes\Psi}(\cdot)$.

If the map Φ *is trace preserving, then the condition of finiteness of the output entropy* $H_\Phi(\cdot)$ *of the map* Φ *can be replaced by the condition*

$$\min\{H_\Phi(\mathrm{tr}_{\mathbb{K}_A}[|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|]), H_\Psi(\mathrm{tr}_{\mathbb{H}_A}[|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|])\} < +\infty, \forall \phi \in \mathbb{H}_A \otimes \mathbb{K}_A.$$

Proof. We may assume that the map Φ is trace nonincreasing.
1. (1) $\Rightarrow$ (2). Let ρ be an arbitrary state in $\mathcal{S}(\mathbb{H}_A) \subset \mathfrak{T}_+(\mathbb{H}_A)$ and ϕ be a vector in $\mathbb{H}_A \otimes \mathbb{K}_A$ such that $\rho = \mathrm{tr}_{\mathbb{K}_A}[|\rho\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|]$. Since the map Ψ is trace preserving, we have $\mathrm{tr}_{\mathbb{K}_A}[(\Phi \otimes \Psi)(|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|)] = \Phi(\rho)$. By noting that $\mathrm{tr}_{\mathbb{H}_A}[(\Phi \otimes \Psi)(|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|)] \leq \Psi(\sigma)$, where $\sigma = \mathrm{tr}_{\mathbb{H}_A}[|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|]$, and by using finiteness of $H(\Psi(\sigma))$ with (7.48) and the triangle inequality
$$\begin{aligned} H((\Phi \otimes \Psi)(|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|)) \geq \big|&H(\mathrm{tr}_{\mathbb{K}_A}[(\Phi \otimes \Psi)(|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|)]) \\ &- H(\mathrm{tr}_{\mathbb{H}_A}[(\Phi \otimes \Psi)(|\phi\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A}\langle\phi|)])\big|, \end{aligned}$$
we conclude that $H(\Phi(\sigma)) < +\infty$. By Theorem 9.2.3, the map Φ has continuous output entropy.
2. (2) $\Rightarrow$ (3) follows from Corollary 9.2.6.
3. (3) $\Rightarrow$ (2) is obvious.

The last assertion of the corollary is proved by the similar argumentation. This proves the corollary. □

9.2.2 Quantum operations

In the following, we follow the approach used in Shirokov [147] to investigate the properties of output entropy $H_\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to [0, +\infty]$, when Φ is a quantum operation, i. e., $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ is a completely positive trace-nonincreasing operator.

Lemma 9.2.8. *Let $(p_i)_{i=1}^{+\infty}$ be a sequence of positive numbers. Then*

$$\sup_{\{x_i\}\subset l_{1+}} H(\{p_i x_i\}_{i=1}^{+\infty}) = \lambda^*$$

where λ^* is either the unique finite solution of the equation $\sum_{i=1}^{+\infty} e^{-\lambda/p_i} = 1$ if it exists or equal to $\lambda^\dagger(\{p_i^{-1}\}_{i=1}^{+\infty}) = \inf\{\lambda > 0 \mid \sum_{i=1}^{+\infty} e^{-\lambda/p_i} < +\infty\}$ otherwise.

In the above, we adopt the convention that $\inf\{\cdots\} = +\infty$ and note that $\sum_{i=1}^{+\infty} e^{-\lambda/p_i} = 1$ has no solution if and only if either $\lambda^\dagger(\{p_i^{-1}\}) = +\infty$ or

$$\sum_{i=1}^{+\infty} \exp(-\lambda^\dagger(\{p_i^{-1}\})/p_i) < 1.$$

Proof. By using the Lagrange method, it is easy to show that the function $\{x_i\}_{i=1}^{+\infty} \mapsto H(\{p_i x_i\}_{i=1}^{+\infty})$ attains its maximum at the vector $(x_i^*)_{i=1}^{+\infty}$ where $x_i^* = c p_i^{-1} e^{-\lambda_n^*/p_i}$ and λ_n^* is the solution of the equation $\sum_{i=1}^{n} e^{-\lambda/p_i} = 1$ and $c = (\sum_{i=1}^{n} p_i^{-1} e^{-\lambda_n^*/p_i})^{-1}$. Hence,

$$\max_{\{x_i\}\subset l_{1+}} H(\{p_i x_i\}_{i=1}^{n}) = \lambda_n^*. \tag{9.12}$$

The assertion of the lemma follows by noting that the sequence $(\lambda_n^*)_{n=1}^{+\infty}$ tends to λ^* as $n \to +\infty$ and by using lower semicontinuity of the classical entropy. This proves the lemma. □

Proposition 9.2.9. *Let $\mathbf{V}$ be a linear operator from $\mathbb{H}_A$ to $\mathbb{H}_B$. The function $\rho \mapsto H(\mathbf{V}\rho\mathbf{V}^*)$, $\rho \in \mathcal{S}(\mathbb{H}_A)$, is continuous if and only if the operator $\mathbf{V}$ is compact and has such sequence $(\nu_i)_{i=1}^{+\infty}$ of eigenvalues of $|\mathbf{V}| := \sqrt{\mathbf{V}^*\mathbf{V}}$ that $\sum_{i=1}^{+\infty} e^{-\lambda/\nu_i^2} < +\infty$ for some $\lambda > 0$. If this condition holds, then*

$$\sup_{\rho\in\mathcal{S}(\mathbb{H}_A)} H(\mathbf{V}\rho\mathbf{V}^*) = \lambda^*(\mathbf{V}),$$

where $\lambda^*(\mathbf{V})$ is either the unique solution of the equation $\sum_{i=1}^{+\infty} e^{-\lambda/\nu_i^2} = 1$ if it exists or equal to $\lambda^\dagger((\nu_i^{-2})_{i=1}^{+\infty}) = \inf\{\lambda > 0 \mid \sum_{i=1}^{+\infty} e^{-\lambda/\nu_i^2} < +\infty\}$, otherwise.

Proof. Without loss of generality, we assume that $\mathbb{H}_A = \mathbb{H}_B = \mathbb{H}$, $\mathbf{V} = |\mathbf{V}|$, $\|\mathbf{V}\|_\infty \leq 1$, and $\ker(\mathbf{V}) = \{\mathbf{0}\}$. Assume that the operator $\mathbf{V}$ takes the form $\mathbf{V} = \sum_{i=1}^{+\infty} v_i |i\rangle_{\mathbb{H}}\langle i|$. If $\sum_{i=1}^{+\infty} e^{-\lambda/v_i^2} < +\infty$ for $\lambda > 0$, then property (4) in Theorem 9.2.3 holds with the basis $\{|i\rangle_{\mathbb{H}}\}_{i=1}^{\infty}$ and the sequence $(h_i)_{i=1}^{+\infty}$, $h_i = \lambda/v_i^2$, since in this case $\Phi^*(\cdot) = \mathbf{V}(\cdot)\mathbf{V}^*$, and hence,

$$\Phi^*(|i\rangle_{\mathbb{H}}\langle i|) = \mathbf{V}(|i\rangle_{\mathbb{H}}\langle i|)\mathbf{V}^*$$
$$= \left(\sum_{j=1}^{+\infty} v_j |j\rangle_{\mathbb{H}}\langle j|\right)(|i\rangle_{\mathbb{H}}\langle i|)\left(\sum_{k=1}^{+\infty} v_k |k\rangle_{\mathbb{H}}\langle k|\right) = v_i^2 |i\rangle_{\mathbb{H}}\langle i|.$$

The assertion concerning the supremum of the function $\rho \mapsto H(\mathbf{V}\rho\mathbf{V}^*)$ is easily derived from Lemma 9.2.8 by using the inequality $H(\mathbf{A}) \leq H(\langle i|\mathbf{A}|i\rangle_{\mathbb{H}})$ for all $\mathbf{A} \in \mathfrak{T}_+(\mathbb{H})$. Suppose the function $\rho \mapsto H(\mathbf{V}\rho\mathbf{V}^*)$ is continuous on the set $\mathcal{S}(\mathbb{H})$. Then the entropy is bounded on the convex set $\{\mathbf{V}\rho\mathbf{V}^* \mid \rho \in \mathcal{S}(\mathbb{H})\}$, and hence, this set is relatively compact by using the construction from the proof of Lemma 9.2.1. Thus, the operator $\mathbf{V}$ is compact (since otherwise there exists a sequence of unit vectors $(|\varphi_n\rangle_{\mathbb{H}})_{n=1}^{+\infty}$ such that the sequence $(\mathbf{V}|\varphi_n\rangle_{\mathbb{H}})_{n=1}^{+\infty}$ is not relatively compact). Lemma 9.2.1 implies existence of an operator $\mathbf{T} \in \mathfrak{T}_1(\mathbb{H})$ such that

$$\sup_{\rho\in\mathcal{S}(\mathbb{H})} \operatorname{tr}[\mathbf{V}\rho\mathbf{V}^*(-\log\mathbf{T})] < +\infty \quad \text{and} \quad \mathbf{UT} = \mathbf{TU}$$

for arbitrary unitary $\mathbf{U}$ commuting with the operator $\mathbf{V}$. It follows from the last property of the operator $\mathbf{T}$ that this operator is diagonizable in the basis $\{|i\rangle_{\mathbb{H}}\}$, i. e., $\mathbf{T} = \sum_{i=1}^{\infty} \tau_i |i\rangle_{\mathbb{H}}\langle i|$, where $(\tau_i)_{i=1}^{+\infty}$ is a sequence of nonnegative numbers such that $\sum_{i=1}^{+\infty} \tau_i \leq 1$. Thus, we have

$$\sup_{\rho\in\mathcal{S}} \operatorname{tr}[\mathbf{V}\rho\mathbf{V}^*(-\log\mathbf{T})] = \sup_{\rho\in\mathcal{S}(\mathbb{H})} \sum_{i=1}^{+\infty} \langle i|\rho|i\rangle_{\mathbb{H}} v_i^2(-\log\tau_i) < +\infty,$$

and hence, $v_i^2(-\log\tau_i) \leq \lambda$ for all i. This implies $\lambda^*(\mathbf{V}) < +\infty$. This proves the proposition. □

Note that an arbitrary quantum operation $\Phi : \mathfrak{T}(\mathbb{H}) \to \mathfrak{T}(\mathbb{H})$ has Kraus representation

$$\Phi(\cdot) = \sum_{i=1}^{+\infty} \mathbf{V}_i(\cdot)\mathbf{V}_i^?$$

where $\{\mathbf{V}_i\}_{i=1}^{+\infty}$ is a set of bounded linear operators from $\mathbb{H}_A$ into $\mathbb{H}_B$ such that $\sum_{i=1}^{\infty} \mathbf{V}_i\mathbf{V}_i^* \leq \mathbf{I}_{\mathbb{H}}$ (correspondingly, $\sum_{i=1}^{+\infty} \mathbf{V}_i\mathbf{V}_i^* = \mathbf{I}_{\mathbb{H}}$ for the quantum channel).

The following proposition contains the sufficient conditions for continuity of the output entropy of a quantum operation Φ expressed in terms of the set $\{\mathbf{V}_i\}_{i=1}^{+\infty}$ of its Kraus operators.

Proposition 9.2.10. *Let Φ be a quantum operation in $\mathfrak{T}_{\leq 1}(\mathbb{H}_A, \mathbb{H}_B)$ and $\{\mathbf{V}_i\}_{i\in I}$ be the corresponding set of Kraus operators. Let $d_i = \operatorname{rank}(\mathbf{V}_i) \leq +\infty$.*

1. *If the index set I is finite, then the operation Φ has continuous output entropy if and only if which in this case is equivalent to*

$$\lambda^*(\mathbf{V}_i) < +\infty, \quad \forall i \in I, \tag{9.13}$$

which in this case is equivalent to

$$\lambda^*\left(\sqrt{\sum_{i\in I} \mathbf{V}_i^* \mathbf{V}_i}\right) < +\infty. \tag{9.14}$$

In the general case, (9.13) *is a necessary condition of continuity of the output entropy of the operation Φ (in contrast to* (9.14)*).*

2. *If $I = \mathbb{N}$, then the operation Φ has continuous output entropy if one of the following conditions is valid:*
   (i) *$d_i < +\infty$ for all i and there exists a sequence $(h_i)_{i=1}^{+\infty}$ of nonnegative numbers such that*

$$\left\| \sum_{i=1}^{+\infty} h_i \mathbf{V}_i^* \mathbf{V}_i \right\|_\infty < +\infty \quad \textit{and} \quad \sum_{i=1}^{+\infty} d_i e^{-h_i} < +\infty$$

   and
   (ii) *$H(\{\operatorname{tr}[\mathbf{V}_i \rho \mathbf{V}_i^* +]\}_{i=1}^{+\infty}) < +\infty$ for all $\rho \in \mathcal{S}(\mathbb{H}_A)$ and condition* (9.14) *holds.*

Note that Condition (9.14) in (ii) can be replaced by the condition of finiteness of one of series $\sum_{i=1}^{+\infty} \lambda^*(\mathbf{V}_i)$ and $\sum_{i=1}^{+\infty} \log(d_i \|\mathbf{V}_i\|_\infty)$. If the sequence $(\mathbf{V}_i)_{i=1}^{+\infty}$ consists of scalar multiples of mutually orthogonal projectors, then (i) is a necessary condition of continuity of the output entropy of the operation Φ. If $\operatorname{range}(\mathbf{V}_i) \perp \operatorname{range}(\mathbf{V}_j)$ for all $i \neq j$, then (ii) is a necessary condition of continuity of the output entropy of the operation Φ.

Proof of Proposition 9.2.10. 1. This directly follows from Corollary 9.2.5, Proposition 9.2.9, since $\sum_{i=1}^{+\infty} \mathbf{V}_i^* \mathbf{V}_i = \Phi^*(\mathbf{I}_{\mathbb{H}_B})$.

2. Suppose condition (i) holds. Let $\mathbb{K} = \bigoplus_{i=1}^{+\infty} \operatorname{range}(\mathbf{V}_i)$ and $\mathbf{U}_i$ be a partial isometry from $\mathbb{H}_B$ into $\mathbb{K}$ such that $\mathbf{U}_i^* \mathbf{U}_i$ is the projector onto $\operatorname{range}(\mathbf{V}_i) \subset \mathbb{H}_B$ and $\mathbf{U}_i \mathbf{U}_i^*$ is the projector onto $\operatorname{range}(\mathbf{V}_i) \subset \mathbb{K}$. Consider the quantum operation $\hat{\Phi}$ in $\mathfrak{T}_{\leq 1}(\mathbb{H}_A, \mathbb{A}_B)$ defined by the sequence of Kraus operators $(\hat{\mathbf{V}}_i)_{i=1}^{+\infty}$, $\hat{\mathbf{V}}_i = \mathbf{U}_i \mathbf{V}_i$. We have $\operatorname{range}(\hat{\mathbf{V}}_i) \perp \operatorname{range}(\hat{\mathbf{V}}_j)$ for all $i \neq j$. Let $\mathbf{P}_i$ be the d_i-rank projector onto the subspace $\operatorname{range}(\hat{\mathbf{V}}_i)$. Consider the $\mathfrak{H}$-operator $\mathbf{H} = \sum_{i=1}^{+\infty} h_i \mathbf{P}_i$. The condition $\sum_{i=1}^{+\infty} d_i e^{-h_i} < +\infty$ means that $\lambda^\dagger(\mathbf{H}) < +\infty$. The condition $\| \sum_{i=1}^{+\infty} h_i \mathbf{V}_i \mathbf{V}_i^* \|_\infty = h < +\infty$ implies

$$\operatorname{tr}[\mathbf{H}\hat{\Phi}(\rho)] = \sum_{i=1}^{+\infty} h_i \operatorname{tr}[\mathbf{P}_i \hat{\mathbf{V}}_i^* \hat{\mathbf{V}}_i] = \operatorname{tr}\left[\sum_{i=1}^{+\infty} h_i \mathbf{V}_i \mathbf{V}_i^* \rho\right] \leq h, \quad \forall \rho \in \mathbb{S}(\mathbb{H}_A).$$

The quantum entropy is bounded on the set $\hat{\Phi}(\mathcal{S}(\mathbb{H}_A))$. Since

$$H(\hat{\Phi}(\rho)) = \sum_{i=1}^{+\infty} H(\mathbf{V}_i \rho \mathbf{V}_i^*) + H\left(\{\operatorname{tr}[\mathbf{V}_i \rho \mathbf{V}_i^*]\}_{i=1}^{+\infty}\right).$$

Inequality (7.49) implies boundedness of the function $\rho \mapsto H(\Phi(\rho))$. By Theorem 9.2.3, this function is continuous. If condition (ii) holds then Proposition 9.2.11 below implies continuity of the output entropy of the operation Φ. Possibility to replace condition (9.14) in (ii) by one of the conditions $\sum_{i=1}^{+\infty} \lambda^*(\mathbf{V}_i) < +\infty$ and $\sum_{i=1}^{+\infty} \log(d_i \|\mathbf{V}_i\|_\infty) < +\infty$ follows from Corollary 9.2.5, since each of these conditions implies finiteness of the series $\sum_{i=1}^{+\infty} H(\mathbf{V}_i \rho \mathbf{V}_i^*)$ for any ρ in $\mathcal{S}(\mathbb{H}_A)$.

To prove the assertion concerning necessity of condition (i), assume that $\Phi(\cdot) = \sum_{i=1} c_i \mathbf{P}_i(\cdot)\mathbf{P}_i$, where $(\mathbf{P}_i)_{i=1}^{+\infty}$ is a sequence of mutually orthogonal projectors. Lemma 9.2.1 implies existence of a trace-class operator of the form $\mathbf{T} = \sum_{i=1}^{+\infty} \lambda_i \mathbf{P}_i$ such that

$$\sup_{\rho \in \mathcal{S}(\mathbb{H}_A)} \operatorname{tr}[\Phi(\rho)(-\log \mathbf{T})] = \sup_{\rho \in \mathcal{S}(\mathbb{H}_A)} \operatorname{tr}\left[\sum_{i=1}^{+\infty} c_i(-\log \lambda_i)\mathbf{P}_i \rho\right] < +\infty.$$

Since $\operatorname{tr}[\mathbf{T}] = \sum_{i=1}^{+\infty} d_i \lambda_i$, condition (i) holds with the sequence $(h_i)_{i=1}^{+\infty}$, $h_i = -\log \lambda_i$.

To prove the assertion concerning necessity of condition (ii), it is sufficient to note that the condition $\operatorname{range}(\mathbf{V}_i) \perp \operatorname{range}(\mathbf{V}_j)$ for all $i \neq j$ implies

$$\begin{aligned} H(\mathbf{V}\rho\mathbf{V}^*) &\leq H(\Phi(\rho)) \\ &= \sum_{i=1}^{+\infty} H(\mathbf{V}_i \rho \mathbf{V}_i^*) + H(\{\operatorname{tr}[\mathbf{V}_i \rho \mathbf{V}_i^*]\}_{i=1}^{+\infty}), \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A), \end{aligned}$$

where $\mathbf{V}$ is the Stinespring contraction of the operation Φ defined via the set $\{\mathbf{V}_i\}_{i=1}^{+\infty}$ (see the proof of Proposition 9.2.11 below), and to apply Theorem 9.2.3 and Proposition 9.2.10 (by using $\mathbf{V}^*\mathbf{V} = \sum_{i=1}^{+\infty} \mathbf{V}_i^* \mathbf{V}_i$). This proves the proposition. □

Example 9.1. Let $(\mathbf{V}_i)_{i=1}^{+\infty}$ be a sequence of finite rank operators in $\mathfrak{B}(\mathbb{H}_A)$ such that $\sum_{i=1}^{+\infty} \mathbf{V}_i^* \mathbf{V}_i \leq \mathbf{I}_{\mathbb{H}_A}$, $\operatorname{range}(\mathbf{V}_i) \perp \operatorname{range}(\mathbf{V}_j)$ for all sufficiently large $i \neq j$ and $\|\mathbf{V}_i\|_\infty^2 \leq C \log^{-\alpha}(i)$ for all i, where $\alpha \geq 0$ and $C > 0$. Since $\mathbf{V}_i^* \mathbf{V}_i \leq C \log^{-\alpha}(i) \mathbf{P}_i$, where $\mathbf{P}_i$ is the projector on the subspace range($\mathbf{V}_i$, condition a) in part 2 of Proposition 9.2.10 holds for the operation $\Phi_\alpha(\cdot) = \sum_{i=1}^{+\infty} \mathbf{V}_i(\cdot)\mathbf{V}_i$ for all $\alpha \geq 1$ provided the rate of increase of the sequence $\left(\operatorname{rank}(\mathbf{V}_i)\right)_{i=1}^{+\infty}$ does not exceed the polynomial rate: $\operatorname{rank}(\mathbf{V}_i) \leq i^n$ for some natural n and all sufficiently large i (this can be shown by using the sequence $(h_i)_{i=1}^{+\infty}$, $h_i = (n+2)\log(i)$. Hence, the output entropy of the operation Φ_α is continuous in this case. The last assertion of Proposition 9.2.10 shows that the output entropy of the operation Φ_α is not continuous if $\alpha < 1$ and $\mathbf{V}_i = \sqrt{C\log^{-\alpha} i}\mathbf{P}_i$ even for the bounded sequence $(\operatorname{rank}(\mathbf{V})_i)_{i=1}^{\infty}$.

The following proposition contains the sufficient conditions for continuity of the output entropy of the complementary operation expressed in terms of the set $(\mathbf{V}_i)_{i=1}^{+\infty}$ of Kraus operators of the initial operation.

Proposition 9.2.11. *Let Φ be a quantum operation in $\mathfrak{T}_{\leq 1}(\mathbb{H}_A, \mathbb{H}_B)$ and $\{\mathbf{V}_i\}_{i=1}^{+\infty}$ be the corresponding set of Kraus operators. The complementary operation $\hat{\Phi}$ has continuous output entropy if one of the following conditions holds:*
1. $H(\{\operatorname{tr}[\mathbf{V}_i\rho\mathbf{V}_i^*]\}_{i=1}^{+\infty}) < +\infty$ *for all* $\rho \in \mathcal{S}(\mathbb{H}_A)$;
2. *there exists a sequence of nonnegative real numbers* $(h_i)_{i=1}^{+\infty}$ *such that*

$$\left\| \sum_{i=1}^{+\infty} h_i \mathbf{V}_i^* \mathbf{V}_i \right\|_\infty < +\infty \quad \textit{and} \quad \sum_{i=1}^{+\infty} e^{-h_i} < +\infty;$$

3. $H(\{\|\mathbf{V}_i\|_\infty\}_{i=1}^{+\infty}) < +\infty$.

If $\operatorname{range}(\mathbf{V}_i) \perp \operatorname{range}(\mathbf{V}_j)$ *for all i ≠ j, then (1)⇔(2) is a necessary condition of continuity of the output entropy of the operation* $\tilde{\Phi}$.

Note that the above conditions are related in the following way: (3) ⇒ (2) ⇔ (1).

Proof. We first show that (1) implies continuity of the function $\rho \mapsto H(\hat{\Phi}(\rho))$. Let $\mathbb{H}_C$ be a separable Hilbert space and $\{|i\rangle_C\}_{i=1}^{+\infty}$ be an orthonormal basis in $\mathbb{H}_C$. Then the operator $\mathbf{V} : \mathbb{H}_A \to \mathbb{H}_B \otimes \mathbb{H}_C$ defined by

$$\mathbf{V}(|\varphi\rangle_A) = \sum_{i=1}^{+\infty} |\mathbf{V}_i\varphi\rangle_{\mathbb{H}_B} \otimes |i\rangle_{\mathbb{H}_C}$$

is the Stinespring contraction for the operation Φ. That is,

$$\Phi(\mathbf{A}) = \operatorname{tr}_B[\mathbf{V}\mathbf{A}\mathbf{V}^*], \quad \mathbf{A} \in \mathfrak{T}(\mathbb{H}_A).$$

Therefore, we have

$$\hat{\Phi}(\mathbf{A}) = \operatorname{tr}_B[\mathbf{V}\mathbf{A}\mathbf{V}^*] = \sum_{i,j=1}^{+\infty} \operatorname{tr}[\mathbf{V}_i\mathbf{A}\mathbf{V}_j^*]|i\rangle_C\langle j|, \quad \mathbf{A} \in \mathfrak{T}(\mathbb{H}_A).$$

By relation $H(\mathbf{A}) = H(\{\langle i|\mathbf{A}|i\rangle_A\}_{i=1}^{+\infty})$, condition (1) implies

$$\begin{aligned} H(\hat{\Phi}(\rho)) &\leq H(\{\langle i|\hat{\Phi}(\rho)|i\rangle_C\}_{i=1}^{+\infty}) \\ &= H(\{\operatorname{tr}[\mathbf{V}_i\rho\mathbf{V}_i^*]\}_{i=1}^{+\infty}) < +\infty, \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A). \end{aligned}$$

By Theorem (9.2.3), the output entropy of the complementary $\hat{\Phi}$ is continuous.

Since finiteness of the function $\rho \mapsto H(\{\mathrm{tr}[\mathbf{V}\rho\mathbf{V}^*]\}_{i=1}^{+\infty})$, $\rho \in \mathcal{S}(\mathbb{H}_A)$, implies that the last condition can be rewritten as follows:

$$\sup_{\rho\in\mathcal{S}(\mathbb{H}_A)} \sum_{i=1}^{+\infty} h_i \,\mathrm{tr}[\mathbf{V}_i\rho\mathbf{V}_i^*] < +\infty$$

and by using the classical versions of part (1) of Proposition 9.2.2 and Lemma 9.2.1. The implication (3) $\Rightarrow$ (2) is obvious.

If rangle$(\mathbf{V}_i) \perp$ range$(\mathbf{V}_j)$ for all $i \neq j$, then

$$\hat{\Phi}(\rho) = \sum_{i=1}^{+\infty} \mathrm{tr}[\mathbf{V}_i\rho\mathbf{V}_i^*]|i\rangle_{\mathbb{H}_C}\langle i|, \quad \forall\rho \in \mathcal{S}(\mathbb{H}_A),$$

and hence, the function in (1) coincides with the output entropy of the operation $\hat{\Phi}$. This proves the proposition. □

9.3 Convex closure of output entropy

We now consider the channel input entropy $H(\cdot)$ and channel output entropy $H_\Phi(\cdot) := H(\Phi(\cdot))$, where $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is a quantum channel from system A to system B.

Definition 9.3.1. Let Φ be a quantum channel or quantum operation from system A to system B. Φ is said to be *preserving finite entropy* (PFE) if

$$H(\rho) < +\infty \Rightarrow H_\Phi(\rho) < +\infty, \tag{9.15}$$

and Φ is said to be *preserving continuous entropy* (PCE) if

$$\lim_{n\to+\infty} H(\rho_n) = H(\rho) \Rightarrow \lim_{n\to+\infty} H_\Phi(\rho_n) = H_\Phi(\rho) \tag{9.16}$$

for all $(\rho_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H}_A)$ such that $\lim_{n\to+\infty} \|\rho_n - \rho\|_1 = 0$.

The following characterization of PFE and PCE channels is due to Shirokov [152, 153] (see also Shirokov and Bulinski [157]).

Theorem 9.3.2 (Shirokov [152, 153]). *Let Φ be a positive linear map from $\mathfrak{T}_+(\mathbb{H}_A)$ into $\mathfrak{T}_+(\mathbb{H}_B)$. The following properties are equivalent:*
(i) *Φ preserves finiteness of the entropy (PFE), i. e., property* (9.15) *holds;*
(ii) *Φ preserves continuity of the entropy, i. e., property* (9.16) *holds;*
(iii) *the output entropy $H_\Phi(\cdot)$ is bounded on* extr$(\mathcal{S}(\mathbb{H}_A))$, *(i. e., the set of all pure states).*

Proof. We may assume that Φ does not increase a trace of any positive operator.

It is clear that (i) $\Rightarrow$ (ii).

By using inequality (7.48) and the spectral decomposition $\rho = \sum_i p_i|\varphi_i\rangle_A\langle\varphi_i|$, we obtain

$$H_\Phi(\rho) \leq \sum_i p_i H_\Phi(|\varphi_i\rangle_A\langle\varphi_i|) + H(\rho) \leq C\|\rho\|_1 + H(\rho), \tag{9.17}$$

where $C = \sup_{\rho\in\mathrm{extr}(\mathcal{S}(\mathbb{H}_A))} H_\Phi(\rho)$. This shows that (iii) ⇒ (i).

To prove the implication (ii) ⇒ (iii). Suppose (iii) was false. Then for each natural n there exists a pure state ρ_n such that $H_\Phi(\rho_n) \geq 2^n$. It follows from (7.48) that the state $\rho_0 = \sum_{n=1}^{+\infty} 2^{-n}\rho_n$ has finite entropy, while the concavity of the function $H_\Phi(\cdot)$ implies $H_\Phi(\rho_0) \geq \sum_{n=1}^{+\infty} 2^{-n} H_\Phi(\rho_n) = +\infty$. Hence (ii) is false. This proves that (ii) ⇒ (iii).

To prove the nontrivial implication (iii) ⇒ (i), it suffices to show that (iii) implies continuity of the function H_Φ on the set $\mathrm{extr}(\mathcal{S}(\mathbb{H}_A))$ and to apply Theorem 9.2.3. This proves the theorem. □

Theorem 9.3.2 implies, in particular, that property (9.15) holds if and only if

$$H^p_{\max}(\Phi) \equiv \sup_{\rho\in\mathrm{extr}(\mathcal{S}(\mathbb{H}_A))} H_\Phi(\rho) < +\infty. \tag{9.18}$$

In analysis of informational properties of a quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$, the convex closure $\overline{\mathrm{co}}(H_\Phi)(\cdot)$ of its output entropy plays important role. Recall from Section 3.4 that the function $\overline{\mathrm{co}}(H_\Phi)(\cdot)$ is defined as the maximal closed (lower semicontinuous) convex function on $\mathcal{S}(\mathbb{H}_A)$ majorized by the function $H_\Phi(\cdot)$. In finite dimensions, $\overline{\mathrm{co}}(H_\Phi)(\cdot)$ coincides with the convex hull $\mathrm{co}(H_\Phi)(\cdot)$ of $H_\Phi(\cdot)$—the maximal convex function on $\mathcal{S}(\mathbb{H}_A)$ majorized by the function $H_\Phi(\cdot)$, which is given by the formula

$$\mathrm{co}(H_\Phi)(\rho) = \inf_{\sum_i p_i\rho_i=\rho} \sum_i p_i H_\Phi(\rho_i), \tag{9.19}$$

where the infimum is over all finite ensembles $\{p_i, \rho_i\}$ of input states with the average state ρ. In infinite dimensions, the function $\overline{\mathrm{co}}(H_\Phi(\cdot))$ coincides with $\mathrm{co}(H_\Phi(\cdot))$ only for positive maps (channels) with finite output entropy, but one can assume that it coincides with the σ-convex hull σ-$co(H_\Phi)(\cdot)$ of $H_\Phi(\cdot)$ defined by formula (9.19) in which the infimum is over all countable ensembles $\{p_i, \rho_i\}$ of input states with the average state ρ. On the other hand, the compactness criterion for families of probability measures on $\mathcal{S}(\mathbb{H}_A)$ makes it possible to show that

$$\overline{\mathrm{co}}(H_\Phi)(\rho) = \inf_{\bar{\rho}(\mu)} \int\limits_{\mathcal{S}(\mathbb{H}_A)} H_\Phi(\rho)\mu(d\rho), \tag{9.20}$$

where the infimum is over all Borel probability measures μ on the set $\mathcal{S}(\mathbb{H}_A)$ with the barycenter ρ. In the following, we show that σ-$\mathrm{co}(H_\Phi(\cdot)) = \overline{\mathrm{co}}(H_\Phi(\cdot))$ under some conditions on Φ.

The following result can be found in Section 3.4 when $f(\cdot) = H_{\Phi}(\cdot) : \mathcal{S}(\mathbb{H}_A) \to [0, +\infty]$.

Proposition 9.3.3. *Let $\Phi : \mathfrak{T}_+(\mathbb{H}_A) \to \mathfrak{T}_+(\mathbb{H}_B)$ be a positive linear map possessing the following (PFE) property:*

$$H(\rho) < +\infty \Rightarrow H_{\Phi}(\rho) < +\infty.$$

Then:
1. $\sigma\text{-co}(H_{\Phi})(\rho) = \overline{\text{co}}(H_{\Phi})(\rho)$ *for any* $\rho \in \mathcal{S}(\mathbb{H}_A)$*;*
2. *the function* $\sigma\text{-co}(H_{\Phi})(\cdot) = \overline{\text{co}}(H_{\Phi})(\cdot)$ *is continuous and bounded on* $\mathcal{S}(\mathbb{H}_A)$*.*

10 Quantum entanglement

As often stated in previous chapters, a closed (or isolated) quantum system is a system that does not interact with any other systems and, therefore, there is no correlation between the systems to speak of. However, an open quantum system is a system that interacts with other systems such as its environment due to a measurement, etc. Therefore, there is strong possibility of correlation of a quantum state on an open quantum system that consists of multiple component systems.

Does quantum correlation exist in open systems? Are quantum correlations different from classical correlations? The answer to these questions is affirmative. Historically, Einstein, Podolsky and Rosen [44] were the first to realize that quantum physics comes with unfamiliar correlations that have no classical interpretation or counterpart. This realization led them to reject quantum mechanics as a whole, despite its remarkable success at that time. In his response to [44], Schrodinger [161] discovered and coined the term *entanglement* for those strange nonclassical correlations.

What is a quantum entanglement? Roughly speaking, a pure quantum state on a multipartite system is said to be an entangled state if it is not separable, i. e., it cannot be written as tensor product of states on the component subsystems. The crucial point about entanglement of states is that this opens the possibility for correlations between subsystems. It turns out that entanglement is an exclusively quantum property and makes it possible for numerous, new promising applications of quantum mechanics in computing, communication and cryptography. In particular, entanglement is of great importance, because it is a central resource in many applications of quantum information theory like entanglement enhanced teleportation or quantum computing.

Due to its interesting applicability, entanglement (see, e. g., [89]) is still one of the most interesting topics in modern physics. This chapter attempts to explore the concepts of entanglement and investigates its relevance in quantum information theory. To explain entanglement in greater detail and to introduce some necessary formalism, we have to complement the scheme developed in previous chapters by a procedure, which allows us to construct states and observables of the composite system from its subsystems by means of tensor products.

10.1 Schmidt decomposability of states

Recall from Definition 2.7.1 that the tensor product $\mathbf{T}_A \otimes \mathbf{T}_B$ of two linear operators $\mathbf{T}_A \in \mathfrak{L}(\mathbb{H}_A)$, $\mathbf{T}_B \in \mathfrak{L}(\mathbb{H}_B)$ is defined first for product vectors $\phi \otimes \psi \in \mathbb{H}_A \otimes \mathbb{H}_B$ by $(\mathbf{T}_A \otimes \mathbf{T}_B)(\psi_1 \otimes \psi_2) = (\mathbf{T}_A \phi) \otimes (\mathbf{T}_B \psi)$ and then it is extended by linearity to all vectors in $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$. In this case, the space $\mathfrak{L}(\mathbb{H}_{AB})$ coincides with the span of all $\mathbf{T}_A \otimes \mathbf{T}_B$. The tensor product $\mathbf{T}_1 \otimes \mathbf{T}_2 \otimes \cdots \otimes \mathbf{T}_n$ of $\mathbf{T}_i \in \mathfrak{L}(\mathbb{H}_i)$ for $i = 1, 2, \ldots, n$ can be similarly defined.

https://doi.org/10.1515/9783110788105-010

Consider the bipartite quantum system AB represented by the Hilbert space $\mathbb{H}_{AB}$. Let $\mathfrak{T}(\mathbb{H}_{AB})$ be the Banach space of trace-class operators under the trace-norm $\|\cdot\|_1$. It can also easily checked that the partial traces $\mathrm{tr}_A[\cdots]$ and $\mathrm{tr}_B[\cdots]$ defined in Subsection 2.8.2, and viewed as maps

$$\mathrm{tr}_A[\cdots] : \mathfrak{T}(\mathbb{H}_{AB}) \to \mathfrak{T}(\mathbb{H}_B)$$

and

$$\mathrm{tr}_B[\cdots] : \mathfrak{T}(\mathbb{H}_{AB}) \to \mathfrak{T}(\mathbb{H}_A)$$

that are completely positive and trace preserving. The partial trace map given above includes a dual map $\mathrm{tr}_B^*[\cdots]$ between the C^*-algebras of bounded operators on $\mathbb{H}_A$ and $\mathbb{H}_{AB}$ given by $\mathrm{tr}_B^*[\mathbf{a}] = \mathbf{a} \otimes \mathbf{I}$. tr_B^* maps observables to observables and is the Heisenberg picture representation of tr_B. The dual map $\mathrm{tr}_A^*[\cdots]$ can similarly represented as $\mathrm{tr}_A^*[\mathbf{b}] = \mathbf{I} \otimes \mathbf{b}$.

Let $\{\phi_i\}_{i=1}^{+\infty}$ and $\{\psi_j\}_{j=1}^{+\infty}$ be orthonormal bases of the Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. As mentioned earlier in this section, $\{\phi_i \otimes \psi_j\}_{i,j=1}^{+\infty}$ forms an orthonormal basis of the Hilbert space $\mathbb{H}_{AB}$ and we can expand each $\zeta \in \mathbb{H}_{AB}$ as $\zeta = \sum_{i,j} \zeta_{ij} \phi_i \otimes \psi_j$ with $\zeta_{ij} = \langle \phi_i \otimes \psi_j, \zeta \rangle_{AB}$, where $\langle \cdot, \cdot \rangle_{AB}$ denotes the inner product on the composite space $\mathbb{H}_{AB}$. This procedure works for an arbitrary number of tensor factors. However, if we have exactly a twofold tensor product, there is a more economic way to expand ζ, called the *Schmidt decomposition* in which only diagonal terms of the form $\{\phi_i \otimes \psi_i\}_{i=1}^{+\infty}$ appear.

Schmidt decomposition of pure states on a tensor product Hilbert space is defined below.

Definition 10.1.1. A pure state ρ_{AB} on the tensor product Hilbert space $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$ is said to have a Schmidt decomposition if there exist an orthonormal basis $\{|i\rangle_A\}_{i=1}^{+\infty}$ of $\mathbb{H}_A$ and an orthonormal basis $\{|i\rangle_B\}_{i=1}^{+\infty}$ of $\mathbb{H}_B$ such that

$$\rho_{AB} = \sum_{i=1}^{+\infty} \sqrt{\lambda_i} |i\rangle_A \otimes |i\rangle_B,$$

for some positive constants $\lambda_i, i = 1, 2, \ldots$. In this case, $\lambda_i, i = 1, 2, \ldots$ are called *Schmidt coefficients* of ρ_{AB} and the number of nonzero $\lambda_i' s$ in this decomposition is called the *Schmidt number*.

A. Finite-dimensional case

To state and prove Schmidt decomposition (in finite dimensions), we first recall (see, e. g., Banerjee and Roy [6]) the singular value decomposition (SVD) as follows. SVD is a factorization of a real or complex matrix. It generalizes the eigendecomposition of a square normal matrix with an orthonormal eigenbasis to any $m \times n$ matrix. It is related to the polar decomposition. Specifically, the singular value decomposition of an $m \times n$

complex matrix M is a factorization of the form $M = U\Sigma V^*$, where U is an $m \times m$ complex unitary matrix, Σ is an $m \times n$ rectangular diagonal matrix with nonnegative real numbers on the diagonal and V is an $n \times n$ complex unitary matrix. If M is real, U and V can also be guaranteed to be real orthogonal matrices. In such contexts, the SVD is often denoted $M = U\Sigma V^\top$. The diagonal entries $\sigma_i = \Sigma_{ii}$ are uniquely determined by M and are known as the singular values of M. The number of nonzero singular values is equal to the rank of M. The columns of U and the columns of V are called left-singular vectors and right-singular vectors of M, respectively. They form two sets of orthonormal bases $\vec{u}_1, \ldots, \vec{u}_m$ and $\vec{v}_1, \ldots, \vec{v}_n$ (treated as column vectors), and the singular value decomposition can be written as $M = \sum_{i=1}^r \sigma_i \vec{u}_i \vec{v}_i^*$, where $r \le \min\{m, n\}$ is the rank of M.

The following Schmidt decomposition theorem of finite-dimensional Hilbert spaces was originally established by Schmidt (see e. g. Wilde [178]). In recent years, it is widely used in quantum information theory (see, e. g., Nielson and Chuang [116], Hayashi [61], Holevo [77], Watrous [173] and Wilde [178]). However, there may not exist a Schmidt decomposition for some pure states on tensor product of infinite-dimensional Hilbert spaces (see Theorem 10.1.6 below).

Lemma 10.1.2. *Let $\mathbb{H}_A$ and $\mathbb{H}_B$ be Hilbert spaces of dimensions n and m, respectively. Assume $n \ge m$. For any vector $\zeta_{AB} \in \mathbb{H}_{AB}$, there exist orthonormal sets $\{\phi_i\}_{i=1}^n \subset \mathbb{H}_A$ and $\{\psi_j\}_{j=1}^m \subset \mathbb{H}_B$ such that $\zeta_{AB} = \sum_{i=1}^m \alpha_i \phi_i \otimes \psi_i$, where the scalars α_i are real, nonnegative and uniquely determined by ζ_{AB}*

Proof. The decomposition claimed in this lemma is essentially a restatement of the singular value decomposition (see the paragraph above) in a different context. Fix orthonormal bases $\{e_i\}_{i=1}^n \subset \mathbb{H}_A$ and $\{f_j\}_{j=1}^m \subset \mathbb{H}_B$. We can identify an elementary tensor $e_i \otimes f_j$ with the $n \times m$ matrix $M = [M_{ij}]$, where $M_{ij} = e_i f_j^\top$ ($f_j^\top$ is the transpose of f_j). A general element of the tensor product $\zeta_{AB} = \sum_{i=i}^n \sum_{j=1}^m \beta_{ij} e_i \otimes f_j$ can then be viewed as the $n \times m$ matrix $M_\zeta = [\beta_{ij}]$. By the singular value decomposition, there exist an $n \times n$ unitary matrix U, $m \times m$ unitary V and a positive semidefinite diagonal $n \times m$ matrix Σ such that

$$M_\zeta = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^*.$$

Write $U = [U_1\ U_2]$, where U_1 is $n \times m$ and we have $M_\zeta = U_1 \Sigma V^*$. Let $\{\phi_1, \ldots, \phi_m\}$ be the m columns vector of U_1, $\{\psi_1, \ldots, \phi_m\}$ be the column vector of V and $\alpha_1, \alpha_2, \ldots, \alpha_m$ the diagonal elements of Σ. The previous expression then becomes $M_\zeta = \sum_{k=1}^m \alpha_k \phi_k \psi_k^*$. Then $\zeta_{AB} = \sum_{k=1}^m \alpha_k \phi_k \otimes \psi_k$. This proves the lemma. □

Proposition 10.1.3 (Finite-dimensional Schmidt decomposition). *Let $\mathbb{H}_A$ and $\mathbb{H}_B$ be two finite-dimensional complex Hilbert spaces with $\dim(\mathbb{H}_A) = \dim(\mathbb{H}_B) = N$. For each $\zeta \in \mathbb{H}_{AB}$, there are orthonormal bases $\{\phi_i\}_{i=1}^N$ and $\{\psi_j\}_{j=1}^N$ of $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively, such that $\zeta = \sum_{i=1}^N \sqrt{\lambda_i} \phi_i \otimes \psi_i$ holds. The two bases $\{\phi_i\}_{i=1}^N$ and $\{\psi_j\}_{j=1}^N$ are uniquely* *determined by* ζ*. The expansion is called Schmidt decomposition and the numbers* $\sqrt{\lambda_i}$ *are the Schmidt coefficients.*

Proof. By definition of pure states (see Definition 2.4.1), it is easy to show that a pure state cannot be written as nontrivial convex combination of another two states by showing one of the coefficients, say $\lambda_{i'} = 1$, and all others $\lambda_i = 0$ for all $i \neq i'$. This implies that any pure state ρ_{AB} on $\mathbb{H}_{AB}$ has the product form

$$\rho_{AB} = (|u_{i'}\rangle_A\langle u_{i'}|) \otimes (|v_{i'}\rangle_B\langle v_{i'}|) \equiv (|\phi\rangle_A\langle\phi|) \otimes (|\psi\rangle_B\langle\psi|),$$

for some $\phi \in \mathbb{H}_A$ and $\psi \in \mathbb{H}_B$.

If ρ_{AB} is mixed, it can be written as a convex combination of pure states

$$\rho_{AB} = \sum_k p_k|\zeta_k\rangle_{AB}\langle\zeta_k|,$$

where $\zeta_k \in \mathbb{H}_{AB}$ and $p_k > 0$ are such that $\sum_k p_k = 1$. Let

$$\rho_{k,A} = \mathrm{tr}_B[|\zeta_k\rangle_{AB}\langle\zeta_k|] \in \mathcal{S}(\mathbb{H}_A) \quad \text{and} \quad \rho_{k,B} = \mathrm{tr}_A[|\zeta\rangle_{AB}\langle\zeta|] \in \mathcal{S}(\mathbb{H}_B).$$

Then

$$\rho_A = \mathrm{tr}_B[\rho_{AB}] = \sum_k p_k\rho_{k,A} \quad \text{and} \quad \rho_B = \mathrm{tr}_A[\rho_{AB}] = \sum_k p_k\rho_{k,B}.$$

But as $\rho_A \in \mathcal{S}(\mathbb{H}_A)$ is a pure state, we only have two possibilities: (i) either $p_{k'} = 1$ for some k' and $p_k = 0$ for $k \neq k'$ or (ii) $\rho_{k,A}$ is the same state ρ_A for every k. In the first case, ρ_A is pure obviously, and for the second option let us introduce the Schmidt decomposition again for each state $|\psi_k\rangle = \sum_{i'}\lambda_{i,k}|u_{i,k}, v_{i,k}\rangle$:

$$\rho_A = \rho_{k,A} = \mathrm{tr}_B[|\zeta_k\rangle_{AB}\langle\zeta_k|] = \sum_i \lambda_{i,k}|u_{i,k}\rangle_A\langle u_{i,k}|, \quad \forall k,$$

so both $\lambda_{i,k}$ and $|u_{i,k}\rangle$ must be independent of k, and

$$\rho_A = \sum_i \lambda_i^2|u_i\rangle\langle u_i|.$$

This equation is the same as the one, where $\lambda_{i'}$ is equal to one for some i', and $\lambda_i = 0$ for all $i \neq i'$. With these requirements, we get $|\psi_k\rangle = |u_{i'}, v_{i',k}\rangle$, and finally

$$\begin{aligned}\rho &= \sum_k p_k|\psi_k\rangle\langle\psi_k| = \sum_k p_k|u_{i'}, v_{i',k}\rangle\langle u_{i'}, v_{i',k}| \\ &= |u_{i'}\rangle\langle u_{i'}| \otimes \sum_k p_k|v_{i',k}\rangle\langle v_{i',k}| \equiv |\psi\rangle_A\langle\psi| \otimes \rho_B.\end{aligned}$$

The proves the proposition. □

B. Infinite-dimensional case

As is shown above, the Schmidt decomposition holds for any pure state on tensor product of any two finite-dimensional Hilbert spaces. However, an example is given by Hu and Yu [91] to show that the Schmidt decomposition theorem is not generally true for bipartite pure states in infinite- dimensional Hilbert spaces.

The following result, due to [91], shows that if a pure state ψ on infinite- dimensional Hilbert space $\mathbb{H}_A \otimes \mathbb{H}_B$ has a Schmidt decomposition then it is necessary that its matrix of amplitude M as defined in Section 10.2.1 has a singular value decomposition.

Proposition 10.1.4. *Let $\mathbb{H}_A$ and $\mathbb{H}_B$ be two infinite-dimensional complex Hilbert spaces. Let ρ_{AB} be a pure state on the composite Hilbert space $\mathbb{H}_{AB}$ and let M be an infinite matrix of amplitude of ρ_{AB}. If ρ_{AB} has a Schmidt decomposition, then there exists unitary operators* **U** *and* **V** *and a diagonal operator* **D** *with nonnegative eigenvalues on $l^2(\mathbb{N};\mathbb{C})$ such that $M = \mathbf{UDV}$.*

Proof. From the definition of the Schmidt decomposition for a bipartite pure state ρ_{AB}, there exist orthonormal bases $\{|i\rangle_A\}_{i=1}^{+\infty}$ for system A and $\{|j\rangle_B\}_{j=1}^{+\infty}$ for system B, such that

$$\rho_{AB} = \sum_{i=1}^{+\infty} \lambda_i |i\rangle_A \otimes |i\rangle_B, \tag{10.1}$$

where λ_i are nonnegative real numbers, which satisfy $\sum_{i=1}^{+\infty} \lambda_i^2 = 1$ known as Schmidt coefficients. Let $M = [a_{ij}]_{i,j=1}^{+\infty}$ be the matrix of the amplitudes of ρ_{AB}, where

$$a_{ij} = \langle \rho_{AB}, |i\rangle_A \otimes |j\rangle_B \rangle_{\mathbb{H}_{AB}}.$$

Since $\{|i\rangle_A\}_{i=1}^{+\infty}$ is an orthonormal basis for Hilbert space $\mathbb{H}_A$, there exists an infinite-dimensional unitary matrix $\mathbf{U} = [u_{ij}]_{i,j=1}^{+\infty}$ such that

$$|i\rangle_A = \sum_{j=1}^{+\infty} u_{ji} |j\rangle_B, \quad i = 1, 2, \ldots. \tag{10.2}$$

Similarly, there exists an unitary matrix $\mathbf{V} = [v_{mn}]_{m,n=1}^{+\infty}$ such that

$$|j\rangle_B = \sum_{i=1}^{+\infty} v_{ij} |i\rangle_A, \quad j = 1, 2, \ldots. \tag{10.3}$$

Substituting (10.2) and (10.3) into (10.1), we have

$$\psi = \sum_{i,j,k=1}^{+\infty} u_{ji} \lambda_i v_{ik} |j\rangle_A \otimes |k\rangle_B. \tag{10.4}$$

Comparing (10.4) with (10.1), we have $a_{jk} = \sum_{i=1}^{+\infty} u_{ji} \lambda_i v_{ik}$. Let

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & \cdots \\ 0 & \lambda_2 & \cdots & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Obviously, $\mathbf{D}$ is an diagonal operator and $M = \mathbf{UDV}$. This proves the theorem. □

Corollary 10.1.5. *Suppose M is the infinite-dimensional matrix of the amplitudes of pure state ψ on $\mathbb{H}_{AB}$, which has Schmidt decomposition. Then there exist a unitary operator $\tilde{\mathbf{U}}$ and a Hermitian operator $\mathbf{P}$ on $l^2(\mathbb{N};\mathbb{C})$ such that $M = \tilde{\mathbf{U}}\mathbf{P}$.*

Proof. From the above Proposition 10.1.4, there exist unitary operators $\mathbf{U}$ and $\mathbf{V}$, such that $M = \mathbf{UDV} = \mathbf{UVV}^*\mathbf{DV}$, where $\mathbf{D}$ is a diagonal operator. Let $\tilde{\mathbf{U}} = \mathbf{UV}$, $\mathbf{P} = \mathbf{V}^*\mathbf{DV}$, then $M = \tilde{\mathbf{U}}\mathbf{P}$. It is easily seen that $\tilde{\mathbf{U}}$ is a unitary operator and $\mathbf{P}$ is a Hermitian operator. This proves the corollary. □

Theorem 10.1.6 (Hu and Yu [91]). *There exists a pure state ψ of the composite systems A and B, which does not have Schmidt decomposition.*

Proof. Let

$$M = [a_{ij}]_{i,j=1}^{+\infty} = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \cdots & 0 & 0 & \cdots \\ 0 & 0 & \frac{1}{\sqrt{4}} & 0 & \cdots & 0 & 0 & \cdots \\ 0 & 0 & 0 & \frac{1}{\sqrt{8}} & \cdots & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{\sqrt{2^n}} & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

It is clear that $\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}|a_{ij}|^2 = 1$. Therefore, M is the matrix of amplitudes for some pure state ψ on the composite system $\mathbb{H}_{AB}$. If ψ has a Schmidt decomposition, then from Corollary 10.1.5 there exist a unitary operator $\tilde{\mathbf{U}}$ and a Hermitian operator $\mathbf{P}$, such that

$$M = \tilde{\mathbf{U}}\mathbf{P} \quad \text{or equivalently } \mathbf{P} = \tilde{\mathbf{U}}^* M.$$

Under the orthonormal basis $|i\rangle_A \otimes |j\rangle_B$, $\tilde{\mathbf{U}}^*$ is represented as an infinite- dimensional matrix

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \cdots \\ a_{21} & a_{22} & \cdots & a_{2n} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Since $\mathbf{P}$ is Hermitian, we have $\mathbf{P} = \mathbf{P}^*$. Thus,

$$\tilde{\mathbf{U}}^* M = M^* \tilde{\mathbf{U}}$$

or equivalently

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \cdots \\ a_{21} & a_{22} & \cdots & a_{2n} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & 0 & \dots \\ 0 & 0 & \frac{1}{\sqrt{4}} & 0 \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \cdots \\ 0 & \frac{1}{\sqrt{4}} & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \overline{a}_{11} & \overline{a}_{21} & \overline{a}_{31} & \cdots \\ \overline{a}_{12} & \overline{a}_{22} & \overline{a}_{32} & \cdots \\ \overline{a}_{13} & \overline{a}_{23} & \overline{a}_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

From the above equality, we see $a_{1j} = 0, 1 \le j < \infty$. Thus,

$$\tilde{\mathbf{U}}^* = \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots \\ a_{21} & a_{22} & \cdots & a_{2n} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

and

$$\tilde{\mathbf{U}} = \begin{pmatrix} 0 & \overline{a}_{21} & \cdots & \overline{a}_{n1} & \cdots \\ 0 & \overline{a}_{22} & \cdots & \overline{a}_{n2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \overline{a}_{2n} & \cdots & \overline{a}_{nn} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Obviously, the vector $e_1 = (1, 0, 0, \ldots)^* \in l^2(\mathbb{N}; \mathbb{C})$ belongs to the kernel of $\tilde{\mathbf{U}}$, which means that $\ker(\tilde{\mathbf{U}}) \neq 0$. This is a contradiction, since $\tilde{\mathbf{U}}$ is a unitary operator on $l^2(\mathbb{N}; \mathbb{C})$. Thus, M cannot be written as the product of a unitary operator and a Hermitian operator. From Corollary 10.1.5, ψ does not have Schmidt decomposition. □

10.2 Separability of states

In this section, we investigate separability of pure states of a composite quantum system in terms of a tensor product of pure states on each of its component subsystems.

10.2.1 Matrix of amplitudes

A. Bipartite pure states

To simplify the illustrations of the concepts of a matrix of amplitudes of a pure state in a composite system, we first consider the composite systems that consist of only two subsystems called *bipartite systems*. The matrix of amplitude of a pure state on composite systems with n subsystems for $n > 2$ will be illustrated in part B. For that purpose, we consider a bipartite quantum system $\mathbb{H} = \mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$ (and often written as $\mathbb{H}_{AB}$), where $\mathbb{H}_A$ and $\mathbb{H}_B$ are separable Hilbert spaces that represent the two quantum subsystems A and B, which may or may not necessarily be the input and output quantum system of a quantum channel discussed in Chapters 8 and 9.

Let $\psi \in \mathcal{S}(\mathbb{H}_{AB})$ be a pure state of the bipartite system AB. That is, ψ cannot be written as a convex combination of any other states on $\mathbb{H}_{AB}$ or equivalently, $\psi \in \operatorname{extr}(\mathcal{S}(\mathbb{H}_{AB}))$ (see Definition 2.4.1 for the definition of pure states).

Let $\{|i\rangle_A\}_{i=1}^{+\infty}$ and $\{|j\rangle_B\}_{j=1}^{+\infty}$ be orthonormal bases for Hilbert space $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. From Subsection 2.7.1, we know $\{|i\rangle_A \otimes |j\rangle_B\}_{i,j=1}^{+\infty}$ is an orthonormal basis for the composite system AB represented by $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$. If ψ is a pure state on $\mathbb{H}_{AB}$, we can give an expansion of ψ under this basis as $\psi = \sum_{i,j=1}^{+\infty} a_{ij}|i\rangle_A \otimes |j\rangle_B$, where

$$a_{ij} := \langle \psi, |i\rangle_A \otimes |j\rangle_B \rangle_{\mathbb{H}_{AB}} \in \mathbb{C}$$

is called an amplitude of ψ with $\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} |a_{ij}|^2 = 1$. Let $M = [a_{ij}]_{i,j=1}^{+\infty}$ be the infinite (but countable, since both $\mathbb{H}_A$ and $\mathbb{H}_B$ are separable Hilbert spaces)-dimensional *matrix of the amplitudes* of pure state ψ on the composite system $\mathbb{H}_{AB}$.

Let $l^2(\mathbb{N}; \mathbb{C})$ be the space of infinite sequences of elements $(x_i)_{i=1}^{+\infty}$ in $\mathbb{C}$ such that $\sum_{i=1}^{+\infty} |x_i|^2 < +\infty$. Recall that (see Example 1.2) that $l^2(\mathbb{N}; \mathbb{C})$ is a complex Hilbert space under the inner product $\langle \cdot, \cdot \rangle_{l^2}$ (or simply written as $\langle \cdot, \cdot \rangle_2$ when there is no danger of ambiguity):

$$\langle x, y \rangle_{l^2} = \sum_{i=1}^{+\infty} x_i \bar{y}_i,$$

$$\forall x = (x_1, x_2, \dots, x_n, \dots), y = (y_1, y_2, \dots, y_n, \dots) \in l^2(\mathbb{N}; \mathbb{C}),$$

where $\overline{y}_n$ denotes the complex conjugate of y_n for $n = 1, 2, \dots$.

In the following, we prove $M = [a_{ij}]_{i,j=1}^{+\infty}$ is a compact linear operator on $l^2(\mathbb{N}; \mathbb{C})$.

Lemma 10.2.1. *Let ψ be a pure quantum state on $\mathbb{H}_{AB}$ and let $M = [a_{ij}]_{i,j=1}^{+\infty}$ be a matrix of amplitude of ψ. Then M is a compact linear operator on the Hilbert space $l^2(\mathbb{N}; \mathbb{C})$.*

Proof. (1) For $n = 1, 2, \dots$, let M_n be the nth column vector of the matrix M, i. e.,

$$M = (M_1, M_2, \dots, M_n, \dots) = \left(M_1^\top, M_2^\top, \dots, M_n^\top, \dots\right)^\top,$$

where $M_n^\top$ denotes the transpose of the column vector M_n. We have for every $x \in l^2(\mathbb{N}; \mathbb{C})$,

$$\begin{aligned} Mx &= \left(M_1^\top, M_2^\top, \dots, M_n^\top, \dots\right)^\top x \\ &= \left(M_1^\top \cdot x, M_2^\top \cdot x, \dots M_n^\top \cdot x, \dots\right)^\top \\ &= \left(\langle M_1, \overline{x}\rangle_2, \langle M_2, \overline{x}\rangle_2, \dots \langle M_n, \overline{x}\rangle_2, \dots\right)^\top, \end{aligned}$$

where $M_i^\top \cdot x = \langle M_i, \overline{x}\rangle_2$ denotes the l_2 inner product of the row vector $M_i^\top$ with x for $i = 1.2. \dots$. Therefore,

$$\begin{aligned} \|Mx\|_2^2 &= \sum_{n=1}^{+\infty} |\langle M_n, \overline{x}\rangle_2|^2 \le \sum_{n=1}^{+\infty} \|M_n\|_2^2 \|x\|_2^2 \\ &= \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} |a_{ij}|^2 \|x\|_2^2 = \|x\|_2^2. \end{aligned}$$

This shows that $M : l^2(\mathbb{N}; \mathbb{C}) \to l^2(\mathbb{N}; \mathbb{C})$ is a bounded linear operator with the operator norm $\|M\|_\infty \le 1$.

(2) To prove M is a compact operator, it suffices to show there is a sequence $(T_n)_{n=1}^{+\infty}$ of operators of finite rank such that $\|M - T_n\|_\infty \to 0$ when $n \to +\infty$. Because $M = [a_{ij}]_{i,j=1}^{+\infty}$ is the matrix of the amplitudes of ψ, we have $\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} |a_{ij}|^2 = 1$. Set the sequence of matrices $\{T_n\}_{n=1}^{+\infty}$ as follows:

$$T_n = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & 0 & 0 & \cdots \\ a_{21} & a_{22} & \dots & a_{2n} & 0 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{10.5}$$

It is clear that the matrix T_n has finite rank n for $n = 1, 2, \ldots$. Denote $s = \sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}|a_{ij}|^2 = 1$ and $s_n = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2$ for $n = 1, 2, \ldots$. It is clear that the sequence $(s_n)_{n=1}^{+\infty}$ converges to s absolutely. We have

$$M - T_n = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_{1(n+1)} & a_{1(n+2)} & \cdots \\ 0 & 0 & \cdots & 0 & a_{2(n+1)} & a_{2(n+2)} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a_{n,n+1} & a_{n,n+2} & \cdots \\ a_{n+1,1} & a_{n+1,2} & \cdots & a_{n+1,n} & a_{n+1,n+1} & a_{n+1,n+2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} = \begin{pmatrix} B_1^\top \\ B_2^\top \\ \vdots \\ B_n^\top \\ B_{n+1}^\top \\ B_{n+2}^\top \\ \vdots \end{pmatrix}.$$

From the fact that $\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}|a_{ij}|^2 = 1$, we have $B_i \in l^2(\mathbb{N};\mathbb{C})$ for all $i = 1, 2, \ldots$. Let $l^2_{\leq 1}(\mathbb{N};\mathbb{C})$ be the closed unit ball of $l^2(\mathbb{N};\mathbb{C})$, i. e.,

$$l^2_{\leq 1}(\mathbb{N};\mathbb{C}) = \{(x_n)_{n=1}^{+\infty} \in l^2(\mathbb{N};\mathbb{C}) \mid \|(x_n)_{n=1}^{+\infty}\|_2 \leq 1\}.$$

We have for every $x \in l^2_{\leq 1}(\mathbb{N};\mathbb{C})$,

$$(M - T_n)x = \begin{pmatrix} B_1^\top \\ B_2^\top \\ \vdots \\ B_n^\top \\ B_{n+1}^\top \\ B_{n+2}^\top \\ \vdots \end{pmatrix} x = \begin{pmatrix} B_1^\top \cdot x \\ B_2^\top \cdot x \\ \vdots \\ B_n^\top \cdot x \\ B_{n+1}^\top \cdot x \\ B_{n+2}^\top \cdot x \\ \vdots \end{pmatrix} = \begin{pmatrix} \langle B_1, \overline{x}\rangle_2 \\ \langle B_2, \overline{x}\rangle_2 \\ \vdots \\ \langle B_n, \overline{x}\rangle_2 \\ \langle B_{n+1}, \overline{x}\rangle_2 \\ \langle B_{n+2}, \overline{x}\rangle_2 \\ \vdots \end{pmatrix}.$$

Then

$$\begin{aligned} \|(M - T_n)x\|_2^2 &= \sum_{i=1}^{+\infty} |\langle B_i, \overline{x}\rangle_2|^2 \\ &\leq \sum_{i=1}^{+\infty} \|B_i\|_2^2 \|x_i\|_2^2 \leq \sum_{i=1}^{+\infty} \|B_i\|_2^2. \end{aligned} \tag{10.6}$$

From the definition of B_i, $i = 1, 2, \ldots$, we have

$$\begin{aligned} \sum_{i=1}^{+\infty} \|B_i\|_2^2 &= \sum_{i=1}^{n} \|B_i\|_2^2 + \sum_{i=n+1}^{+\infty} \|B_i\|_2^2 \\ &= \sum_{i=1}^{n}\sum_{j=n+1}^{+\infty} |a_{ij}|^2 + \sum_{i=n+1}^{+\infty}\sum_{j=1}^{+\infty} |a_{ij}|^2 \\ &= |s - s_n|. \end{aligned} \tag{10.7}$$

From (10.6) and (10.7), we have $\|(M - T_n)x\|_2^2 \le |s - s_n|$ for all $x \in l^2_{\le 1}(\mathbb{N};\mathbb{C})$. From the definition of the norm of the operators on $l^2(\mathbb{N};\mathbb{C})$, we have

$$\|M - T_n\|_\infty^2 = \sup_{x\in l^2_{\le 1}(\mathbb{N};\mathbb{C})} \|(M - T_n)x\|_2^2 \le |s - s_n| \to 0 \quad \text{as } n \to +\infty.$$

This shows that $M : l^2(\mathbb{N};\mathbb{C}) \to l^2(\mathbb{N};\mathbb{C})$ is a compact operator. □

Recall from the polar decomposition theorem 1.8.11 that if $\mathbf{T}$ is a compact operator on Hilbert space $\mathbb{H}$ and if $\mathbf{A}$ is the unique positive square root of $\mathbf{T}^*\mathbf{T}$, then (a) $\|\mathbf{A}h\|_{\mathbb{H}} = \|\mathbf{T}h\|_{\mathbb{H}}$ for all $h \in \mathbb{H}$; and (b) there is a unique operator $\mathbf{U}$ such that $\|\mathbf{U}h\|_{\mathbb{H}} = \|h\|_{\mathbb{H}}$ when $h \perp \ker(\mathbf{T})$, and $\mathbf{U}h = 0$, when $h \in \ker(\mathbf{T})$ and $\mathbf{UA} = \mathbf{T}$.

The following result is due originally to Hu and Yu [91, 92].

Theorem 10.2.2. *Let ψ be a pure state in $\mathbb{H}_{AB}$, and let M be the matrix of the amplitudes of ψ. Then M has polar decomposition.*

Proof. This follows from Lemma 10.2.1 and polar decomposition restated above (see also Theorem 1.8.11). □

Lemma 10.2.3. *If M is the matrix of the amplitudes of a pure state ψ on $\mathbb{H}_{AB}$, then $M = xy^*$, where $x, y \in l^2(\mathbb{N};\mathbb{C})$ if and only if M is a bounded linear operator with rank 1.*

Proof. (⇒) Assume that $M = xy^*$, for some $x, y \in l^2(\mathbb{N};\mathbb{C})$. For every $z \in l^2(\mathbb{N};\mathbb{C})$, write

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \\ \vdots \end{pmatrix}.$$

We have

$$Mz = xy^\dagger z = x(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n, \dots)\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \\ \vdots \end{pmatrix} = x\left(\sum_{i=1}^{\infty} \bar{y}_i z_i\right).$$

This implies that $\text{range}(M) \subset \text{span}(x)$. But $M \neq 0$. Therefore, $\text{rank}(M) = 1$.

(⇐) Assume that $\text{rank}(M) = 1$. Denote $M = (M_1, M_2, \dots, M_i, \dots)$ and without loss of generality, we can assume that $M_1 \neq 0$. Suppose the vector $e_i \in l^2(\mathbb{N};\mathbb{C})$ is the vector with all 0s except for a 1 in the ith coordinate. We have $Me_1 = M_1$ and $Me_i = M_i$. Then

$M_1, M_i \in \text{range}(M)$, but $\text{rank}(M) = 1$, so there exists $\lambda_i \in \mathbb{C}$ such that $M_i = \lambda_i i M_1$. We have

$$\begin{aligned} M &= (M_1, M_2, \dots, M_i, \dots) \\ &= (M_1, \lambda_2 M_1, \dots, \lambda_i \mathcal{M}_1, \dots) \\ &= M_1(1, \lambda_2, \dots, \lambda_i, \dots). \end{aligned}$$

Denote $x = M_1$ and $y = (1, \lambda_2, \dots, \lambda_i, \dots)^*$. This shows that $M = xy^*$. Since M is an amplitude of the pure state ψ, we know $x, y \in l^2(\mathbb{N}; \mathbb{C})$. This proves the lemma. □

B. Multipartite pure states

Here, we extend the concept of matrix of amplitudes for bipartite pure states to pure states on a composite quantum system that consists of n subsystems. Let $\mathbb{H}^{(k)}$ be the complex Hilbert space that represents subsystem A_k for $k = 1, 2, \dots, n$. Then the composite system can be represented by the Hilbert space $\mathbb{H} = \mathbb{H}^{(1)} \otimes \mathbb{H}^{(2)} \otimes \cdots \otimes \mathbb{H}^{(n)}$, where $n \geq 3$ (Note that we have temporally changed the notation of the Hilbert space for the component subsystems from the previously used notation $\mathbb{H}_k$ to $\mathbb{H}^{(k)}$ in order to accommodate double indices used below). If $\{|e_{k_i}^{(k)}\rangle\}_{k_i=1}^{+\infty}$ is a an orthonormal basis for $\mathbb{H}^{(k)}$ for $k = 1, 2, \dots, n$, then it can be shown that

$$\{|e_{1_i}^{(1)}\rangle \otimes |e_{2_i}^{(2)}\rangle \otimes \cdots \otimes |e_{n_i}^{(n)}\rangle \otimes \cdots\}_{k_i=1}^{+\infty}$$

Assume that each of the component Hilbert space is finite-dimensional with $\dim(\mathbb{H}^{(k)}) = d_k$ for $k = 1, 2, \dots, n$. Then any n-partite pure state ψ on a finite-dimensional composite system $\mathbb{H}$ can be expressed as

$$\psi = \sum_{i_1=1}^{d_1} \sum_{i_2=1}^{d_2} \cdots \sum_{i_n=1}^{d_n} a_{i_1 i_2 \cdots i_n} e_{i_1}^{(1)} \otimes e_{i_2}^{(2)} \otimes \cdots \otimes e_{i_n}^{(n)},$$

where $a_{i_1 i_2 \cdots i_n}$ are called the amplitudes of ψ with

$$\sum_{i_1=1}^{d_1} \sum_{i_2=1}^{d_2} \cdots \sum_{i_n=1}^{d_n} |a_{i_1 i_2 \cdots i_n}|^2 = 1.$$

Consider a n-partite composite system $A_1 A_2 \cdots A_n$ represented by the Hilbert space

$$\mathbb{H}_{A_1 A_2 \cdots A_n} = \mathbb{H}_{A_1} \otimes \mathbb{H}_{A_2} \otimes \cdots \otimes \mathbb{H}_{A_n}.$$

We have the following result for the von Neumann entropy for the n-partite system.

Lemma 10.2.4. *Let $\omega_{A_1A_2\cdots A_n}$ be a quantum state of an n-partite system $A_1A_2\cdots A_n$ and $(P^k_{A_i})_{k=1}^{+\infty} \subset \mathfrak{B}(\mathbb{H}_{A_i})$ be sequences of projectors strongly converging to the identity operators $\mathbf{I}_{A_i}$ for $i = 1, 2, \ldots n$. Let*

$$\omega^k_{A_1A_2\ldots A_n} = \lambda_k^{-1}\mathbf{Q}_k\omega_{A_1A_2,\ldots,A_n}\mathbf{Q}_k,$$

where $\mathbf{Q}_k = \mathbf{P}^k_{A_1} \otimes \mathbf{P}^k_{A_2} \otimes \cdots \otimes \mathbf{P}^k_{A_n}$, $\lambda_k = \mathrm{tr}[\mathbf{Q}_k\omega_{A_1A_2\ldots A_n}]$ and $A_{i_1}\cdots A_{i_m}$, $(m \le n)$, be a subsystem of $A_1\ldots A_n$. Then

$$\lim_{k\to+\infty} H(\omega^k_{A_{i_1}A_{i_2}\cdots A_{i_m}}) = H(\omega_{A_{i_1}\ldots A_{i_m}}) \le +\infty.$$

Proof. By noting that

$$\lambda_k\omega^k_{A_{i_1}A_{i_2}\cdots A_{i_m}} \le (\mathbf{P}^k_{A_{i_1}} \otimes \mathbf{P}^k_{A_{i_2}} \otimes \cdots \otimes \mathbf{P}^k_{A_{i_m}})\omega(\mathbf{P}^k_{A_{i_1}} \otimes \mathbf{P}^k_{A_{i_2}} \otimes \cdots \otimes \mathbf{P}^k_{A_{i_m}}),$$

this assertion can be proved by using Simon's convergence theorems for the von Neumann entropy (see the Appendix of Lieb and Ruskai [105]). This proves the lemma. □

10.2.2 Separability

Roughly speaking, we call a state ρ_{AB} of an (infinite-dimensional as well as finite-dimensional) bipartite quantum systems AB entangled if it cannot be written as a mixture of product states. More precisely, we have the following.

Definition 10.2.5. A quantum state ρ_{AB} on $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$ is said to be *separable* if ρ_{AB} is in the convex closure of the subset of all tensor product states $\omega_A \otimes \sigma_B$, where $\omega_A \in \mathcal{S}(\mathbb{H}_A)$ and $\sigma_B \in \mathcal{S}(\mathbb{H}_B)$. A state is said to be *entangled* if it is not separable.

Since any quantum state in each of the component systems can be written as a convex combination of pure states in that system, therefore, the above definition can be restated as follows: A state $\rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$ is separable if it is in the convex closure of the subset of all tensor products $\rho_A \otimes \rho_B$ of pure states. Note that the convex subset of pure states in $\mathcal{S}(\mathbb{H}_{AB})$ is closed in the trace-norm topology.

The separability of multipartite state $\rho_{1,2,\ldots,n}$ in $\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2 \otimes \cdots \otimes \mathbb{H}_n)$ can be defined similarly as follows.

Definition 10.2.6. A multipartite state $\rho_{1,2,\ldots,n}$ in $\mathcal{S}(\mathbb{H}_1 \otimes \mathbb{H}_2 \otimes \cdots \otimes \mathbb{H}_n)$ is said to be separable if it is in the convex closure of the set of all tensor product states $\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$, where ρ_i is a pure state in $\mathcal{S}(\mathbb{H}_i)$ for $i = 1, 2, \ldots, n$. The state $\rho_{1,2,\ldots,n}$ is said be an entangled state if it is not a separable state.

A. Bipartite quantum states

Let $\psi \in \mathcal{S}(\mathbb{H}_{AB})$ be a quantum state of a bipartite system AB. The following lemma characterizes separability of ψ in terms of its integral representation.

Lemma 10.2.7. *A quantum state $\rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$ is separable if and only if there exists a Borel probability measure μ on $\mathcal{S}(\mathbb{H}_A) \times \mathcal{S}(\mathbb{H}_B)$ such that*

$$\rho_{AB} = \int\limits_{\mathcal{S}(\mathbb{H}_A)} \int\limits_{\mathcal{S}(\mathbb{H}_B)} (|\varphi\rangle_A\langle\varphi| \otimes |\psi\rangle_B\langle\psi|)\mu(d\varphi d\psi), \tag{10.8}$$

where $|\varphi\rangle_A\langle\varphi| \otimes |\psi\rangle_B \otimes \langle\psi| = \langle|\varphi\rangle_{\mathbb{H}_A}\langle\varphi| \otimes |\psi\rangle_{\mathbb{H}_B}\langle\psi|$.

Proof. Let $\mathcal{A}$ be the convex hull of a tensor product of pure states $|\psi\rangle_A\langle\psi|$ in $\mathcal{S}(\mathbb{H}_A)$ with pure states $|\phi\rangle_B\langle\phi|$ in $\mathcal{S}(\mathbb{H}_B)$. That is,

$$\mathcal{A} = \operatorname{co}(\{|\phi\rangle_A\langle\phi| \otimes |\psi\rangle_B\langle\psi| \mid \phi \in \mathbb{H}_A, \psi \in \mathbb{H}_B\}).$$

It is clear that $\mathcal{A}$ is a closed subset of $\mathcal{S}(\mathbb{H}_{AB})$. Lemma 3.3.7 states that $\overline{\operatorname{co}}(\mathcal{A})$ coincides with the set of barycenters of all probability measures supported by $\mathcal{A}$. From Definition 10.2.5, pure state $\rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$ is separable if and only if $\rho_{AB} \in \overline{\operatorname{co}}(\mathcal{A})$, and hence ρ_{AB} is barycenter of a Borel probability measure μ supported by $\mathcal{A}$ or equivalently by $\mathcal{S}(\mathbb{H}_A) \otimes \mathcal{S}(\mathbb{H}_B)$. Therefore, ρ_{AB} is separable if and only if

$$\rho_{AB} = \int\limits_{\mathcal{S}(\mathbb{H}_A)} \int\limits_{\mathcal{S}(\mathbb{H}_B)} (|\varphi\rangle_A\langle\varphi| \otimes |\psi\rangle_B\langle\psi|)\mu(d\varphi d\psi).$$

This proves the lemma. □

Definition 10.2.8. A separable bipartite state $\rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$ is said to be countably decomposable if the measure μ in its integral representation (10.8) is atomic. That is, if there exists a discrete ensemble

$\{p_{ij}, |\phi_i\rangle_A\langle\phi_i| \otimes |\psi_j\rangle_B\langle\psi_j|\}_{i,j=1}^{+\infty}$ such that

$$\rho_{AB} = \sum_{i,j=1}^{+\infty} p_{ij}(|\phi_i\rangle_A\langle\phi_i| \otimes |\psi_i\rangle_B\langle\psi_j|), \tag{10.9}$$

where $p_{ij} > 0$ with $\sum_{i,j=1}^{+\infty} p_{ij} = 1$ and $|\phi_i\rangle_A\langle\phi_i| \otimes |\psi_j\rangle_B\langle\psi_j| \in \mathcal{S}(\mathbb{H}_A) \otimes \mathcal{S}(\mathbb{H}_B)$.

The following lemma follows immediately from the above definition.

Lemma 10.2.9. *If the bipartite state ρ_{AB} is countably decomposable, then there exist $|\alpha\rangle_A \in \mathbb{H}_A$ and $|\beta\rangle_B \in \mathbb{H}_B$ such that*

$$\rho_{AB} \geq |\alpha\rangle_A\langle\alpha| \otimes |\beta\rangle_B\langle\beta|.$$

We assume here that $\mathbb{H}_A$ and $\mathbb{H}_B$ both have a countable number of dimensions (finite or infinite dimensions). That is, $\mathbb{H}_A$ and $\mathbb{H}_B$ both have countable dense subsets or equivalently both $\mathbb{H}_A$ and $\mathbb{H}_B$ are separable Hilbert spaces. Let $\{|i\rangle_A\}_{i=1}^{+\infty}$ and $\{|j\rangle_B\}_{j=1}^{+\infty}$ be orthonormal bases for Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. From above, we know $\{|i\rangle_A \otimes |j\rangle_B\}_{i,j=1}^{+\infty}$ is an orthonormal basis for the tensor product Hilbert space $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$. Let $M = [a_{ij}]_{i,j=1}^{+\infty}$ be the infinite (but countable)-dimensional matrix of amplitudes of the pure sate ψ on $\mathbb{H}_{AB}$, where

$$a_{ij} := \langle \psi, |i\rangle_A \otimes |j\rangle_B \rangle_{\mathbb{H}_{AB}} \in \mathbb{C}$$

with $\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} |a_{ij}|^2 = 1$.

We have proved M is a compact linear operator on $l^2(\mathbb{N};\mathbb{C})$ (see Lemma 10.2.1) and it has a polar decomposition (see Theorem 1.8.11 for polar decomposition of a compact operator).

The following result extends separability of bipartite state from finite-dimensional Hilbert spaces to infinite-dimensional ones is due to Hu and Yu [91].

Theorem 10.2.10. *A pure state ψ on $\mathbb{H}_{AB}$ is separable if and only if there exist two unit vectors $x, y \in l^2(\mathbb{N};\mathbb{C})$ such that $M = xy^*$, where M is the matrix of amplitudes of ψ.*

Proof. ($\Rightarrow$) Assume that the pure state ψ on $\mathbb{H}_{AB}$ is separable. Then ψ can be written as tensor product of a pure state $\phi_A = \sum_{i'=1}^{+\infty} x_{i'} |i'\rangle_A$ on $\mathbb{H}_A$ and a pure state $\phi_B = \sum_{j'=1}^{+\infty} y_{j'} |j'\rangle_B$ on $\mathbb{H}_B$, where $\{|i'\rangle_A\}_{i'=1}^{+\infty}$ and $\{|j'\rangle_B\}_{j'=1}^{+\infty}$ are some orthonormal bases of $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively. That is,

$$\psi = \left(\sum_{i'=1}^{+\infty} x_{i'} |i'\rangle_A \right) \otimes \left(\sum_{j'=1}^{+\infty} y_{j'} |j'\rangle_B \right),$$

where $x = (x_{i'})_{i'=1}^{+\infty}$, $y = (y_{j'})_{j'=1}^{+\infty} \in l^2(\mathbb{N};\mathbb{C})$ are such that $\sum_{i'=1}^{+\infty} |x_{i'}|^2 = 1$, $\sum_{j'=1}^{+\infty} |y_{j'}|^2 = 1$. As above, $\psi = \sum_{i,j=1}^{+\infty} a_{ij} |i\rangle_A \otimes |j\rangle_B$ is the Fourier expansion of ψ under the orthonormal basis $\{|i\rangle_A \otimes |j\rangle_B\}_{i,j=1}^{+\infty}$ in $\mathbb{H}_{AB}$. From the definition of tensor product for infinite-dimensional Hilbert spaces, the amplitudes a_{ij}, $i,j = 1,2,\ldots$ of the pure state ψ on $\mathbb{H}_{AB}$ can be computed as

$$\begin{aligned} a_{ij} &= \langle \psi, |i\rangle_A \otimes |j\rangle_B \rangle_{AB} \\ &= \left\langle \left(\sum_{i'=1}^{+\infty} x_{i'} |i'\rangle_A \right) \otimes \left(\sum_{j'=1}^{+\infty} y_{j'} |j'\rangle_B \right), |i\rangle_A \otimes |j\rangle_B \right\rangle_{AB} \\ &= \left\langle \sum_{i'=1}^{+\infty} x_{i'} |i'\rangle_A, |i\rangle_A \right\rangle_A \left\langle \sum_{j'=1}^{+\infty} y_{j'} |j'\rangle_B, |j\rangle_B \right\rangle_B \\ &= \langle x_i |i\rangle_A, |i\rangle_A \rangle_A \langle y_j |j\rangle_B, |j\rangle_B \rangle_B = x_i y_j, \quad i,j = 1,2,\ldots, \end{aligned}$$

where $\langle \cdot,\cdot \rangle_A = \langle \cdot,\cdot \rangle_{\mathbb{H}_A}$, etc.

Set

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \\ \vdots \end{pmatrix}.$$

From the fact that $\sum_{i=1}^{+\infty} |x_i|^2 = 1$ and $\sum_{j=1}^{+\infty} |y_j|^2 = 1$, we have $x, y \in l^2(\mathbb{N}; \mathbb{C})$ and

$$M = [a_{ij}]_{i,j=1}^{+\infty} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \cdots \\ a_{21} & a_{22} & \cdots & a_{2n} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$
$$= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \end{pmatrix} (y_1, y_2, \ldots, y_j, \ldots) = xy^*.$$

($\Leftarrow$) Suppose that $M = xy^*$, for some $x, y \in l^2(\mathbb{N}; \mathbb{C})$, where $\|x\|_2^2 := \sum_{i=1}^{+\infty} |x_i|^2 < \infty$ and $\|y\|_2^2 := \sum_{j=1}^{+\infty} |y_j|^2 < +\infty$. Because $M = [a_{ij}]_{i,j=1}^{+\infty}$ is the matrix of the amplitudes of the pure state ψ of $\mathbb{H}_A \otimes \mathbb{H}_B$, we know $\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} |a_{ij}|^2 = 1$. Since $M = xy^*$, we see $a_{ij} = x_i \bar{y}_j$ from the multiplication for infinite (countable)-dimensional matrices. Also from the fact that $\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} |a_{ij}|^2 = 1$, we have $\sum_{i=1}^{+\infty} |x_i|^2 \sum_{j=1}^{+\infty} |y_j|^2 = 1$. Setting $\tilde{x} = \frac{x}{\|x\|_2}$ and $\tilde{y} = \frac{y}{\|y\|_2}$, we also have $M = \tilde{x}\tilde{y}$. Therefore, without loss of generality, we can assume $\|x\|_2 = 1$ and $\|y\|_2 = 1$ and construct two vectors $|v\rangle_A = \sum_{i=1}^{+\infty} x_i |i\rangle_A$ and $|w\rangle_B = \sum_{j=1}^{+\infty} \bar{y}_j |j\rangle_B$. We have

$$\begin{aligned} |v\rangle_A \otimes |w\rangle_B &= \left(\sum_{i=1}^{+\infty} x_i |i\rangle_A \right) \otimes \left(\sum_{j=1}^{+\infty} \bar{y}_j |j\rangle_B \right) \\ &= \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} x_i \bar{y}_j (|i\rangle_A \otimes |j\rangle_B) = \sum_{i,j} a_{ij} (|i\rangle_A \otimes |j\rangle_B) = \psi. \end{aligned}$$

This shows that ψ is a separable pure state on $\mathbb{H}_A \otimes \mathbb{H}_B$. This proves the theorem. □

Lemma 10.2.11. *If M is the matrix of the amplitudes of a pure state ψ, then $M = xy^*$ if and only if the determinants of all the 2×2 submatrices of M are zero.*

Proof. (⇒) Assume that $M = xy^*$, where $M = [a_{ij}]_{i,j=1}^{+\infty}$, $x = (x_i)_{i=1}^{+\infty}$, and $y = (y_j)_{j=1}^{+\infty}$. As above, we see $a_{ij} = x_i\bar{y}_j$ for all $i,j = 1,2,\ldots$. Let

$$M_{2\times 2} = \begin{pmatrix} a_{il} & a_{ik} \\ a_{jl} & a_{jk} \end{pmatrix}$$

be any 2×2 submatrix of M. It is easy to check that

$$\det(M_{2\times 2}) = a_{il}a_{jk} - a_{ik}a_{jl} = x_i\bar{y}_l x_j\bar{y}_k - x_i\bar{y}_k x_j\bar{y}_l = 0.$$

Therefore, if the pure state ψ is separable, then the determinants of all the 2×2 submatrices of M are zero.

(⇐) Assume that the determinant of all 2×2 submatrices of M are zeros. Suppose the matrix M can be written as $M = (M_1, M_2, \ldots, M_j, \ldots)$ where $M_1 \neq 0$. If M_1 and M_j are linearly independent for some $j > 1$, then any 2×2 submatrix of (M_1, M_j) is invertible and, therefore, its determinant is zero. This is a contradiction. Therefore, M_1 and M_j are linearly dependent for all $j > 1$. This implies that for each $j > 1$ there exists a nonzero $\lambda_j \in \mathbb{C}$ such that $M_j = \lambda_j M_1$. Then

$$\begin{aligned} M &= (M_1, M_2, \ldots, M_j, \ldots) \\ &= (M_1, \lambda_2 M_1, \ldots, \lambda_j M_1, \ldots) \\ &= M_1(1, \lambda_2, \ldots, \lambda_j, \ldots) = xy^*, \end{aligned}$$

where $x = M_1$ and $y = (1, \bar{\lambda}_2, \ldots, \bar{\lambda}_j, \ldots)^\top$. Since M is the matrix of amplitudes of the pure state ψ, it is clear that $x, y \in l^2(\mathbb{N}; \mathbb{C})$. This proves the lemma. □

The following theorem follows from Theorem 10.2.10 and Lemma 10.2.11.

Theorem 10.2.12. *ψ is a separable pure state $\mathbb{H}_{AB}$ if and only if the determinate of all the 2×2 submatrices of M are zero.*

B. Multipartite quantum states

Let ψ be a pure state on an n-partite tensor product Hilbert space

$$\mathbb{H} = \mathbb{H}^{(1)} \otimes \mathbb{H}^{(2)} \otimes \cdots \otimes \mathbb{H}^{(n)}.$$

We assume that all component subsystems are finite-dimensional and denote $d_k = \dim(\mathbb{H}^{(k)})$ for $k = 1,2,\ldots,n$ and $d = \max\{d_1, d_2, \ldots, d_n\}$. Let $\{e_{k_i}^{(k)}\}_{k_i=1}^{d_k}$.

Let ψ be a pure state on composite system $\mathbb{H}$. The pure state ψ is said to be separable if it can be written as tensor product of pure states on its subsystem, i. e., $\psi = \psi^{(1)} \otimes \psi^{(2)} \otimes \cdots \otimes \psi^{(n)}$, where $\psi^{(k)}$ is a pure state on $\psi^{(k)}$ for $k = 1,2,\ldots,n$.

We have the following result (due to Hu and Yu [91]).

Theorem 10.2.13. *A n-partite pure state ψ is separable if and only if it is Schmidt decomposable and has Schmidt number 1.*

Proof. ($\Rightarrow$) Assume that the pure state ψ is separable. Then $\psi = \psi^{(1)} \otimes \psi^{(2)} \otimes \cdots \otimes \psi^{(n)}$, where $\psi^{(k)}$ is a pure state on $\mathbb{H}^{(k)}$ for $k = 1, 2, \ldots, n$. We construct an orthonormal basis $\{|e_{k_i}^{(k)}\rangle\}_{k_i=1}^{d_k}$ for $k = 1, 2, \ldots, n$ as follows. First, for $k = 1, 2, \ldots, n$, let $|e_1^{(k)}\rangle = \psi^{(k)}$. Then we choose $|e_{i_k}^{(k)}, i_k = 2, 3, \ldots, n$ so that $\{|e_{i_k}^{(k)}\}_{i_k=1}^{d_k}$ forms an orthonormal basis for $\mathbb{H}^{(k)}$ for $k = 1, 2, \ldots, n$. Now choose $\lambda_1 = 1$ and $\lambda_2 = \lambda_3 = \cdots = \lambda_d = 0$. Then the n-partite pure state ψ can be written as

$$\psi = \sum_{k=1}^{d} \lambda_k |e_k^{(1)}\rangle \otimes |e_k^{(2)}\rangle \otimes \cdots |e_k^{(n)}\rangle,$$

which is a Schmidt decomposition with the Schmidt number 1.

($\Leftarrow$) We assume that the n-partite pure state ψ is Schmidt decomposable and has Schmidt number 1. Assume that $\lambda_1 \neq 0$ and we have

$$\begin{aligned}\psi &= \sum_{k=1}^{d} \lambda_k |e_k^{(1)}\rangle \otimes |e_k^{(2)}\rangle \otimes \cdots \otimes |e_k^{(n)}\rangle \\ &= \lambda_k |e_k^{(1)}\rangle \otimes |e_k^{(2)}\rangle \otimes \cdots \otimes |e_k^{(n)}\rangle,\end{aligned}$$

where $|e_1^{(k)}\rangle$ is a pure state on $\mathbb{H}^{(k)}$ for $k = 1, 2, \ldots, n$. This proves the theorem. □

Taking into account of Theorem 10.2.13 separability of any n-partite pure state can be proved by finding its Schmidt decomposition. In order to obtain the Schmidt decomposition of a n-partite pure state ψ, we need to compute (1) the density operator ρ_ψ, (2) the reduced density operators $\rho_\psi^{(k)}$ of ρ_ψ for $k = 1, 2, \ldots, n$ and (3) the eigenvalues λ_{i_k} of $\rho_\psi^{(k)}$ for $k = 1, 2, \ldots, n$. However, it is hard to compute all the eigenvalues of high-dimensional density operators exactly. Theorem 10.2.13 does not give us an effective criterion. In the following, we will introduce some effective separability criterions. Any quantum state can be represented by a density operator. For an n-partite pure state ψ on the composite system $\mathbb{H}$, its density operator is $\rho_\psi = |\psi\rangle\langle\psi|$. Taking the partial trace operation on each kth subsystem $\mathbb{H}^{(k)}$, we get reduced density $\rho_\psi^{(k)} = \mathrm{tr}_{\mathbb{H}^{(k)}}(\rho_\psi)$.

For $k = 1, 2, \ldots, n$, where $M_\psi^{(k)}$ are matrices related to the reduced density operators $\rho_\psi^{(k)}$ and E_k are the identity matrices with the same dimension as $M_\psi^{(k)}$ for $k = 1, 2, \ldots, n$.

The following lemma can be easily proved by using the characterizations of density operators and the proof is omitted here.

Lemma 10.2.14. *The rank of $M_\psi^{(k)}, k = 1, 2, \ldots, n$ are all equal to 1 if and only if the following condition holds:*

$$\det(M_\psi^{(k)} - E_k) = 0, \quad k = 1, 2, \ldots, n. \tag{10.10}$$

In order to prove the separability of an n-partite pure state ψ, there is no need to find the Schmidt decomposition because of the following result, which is also due to Hu and Yu [91].

Theorem 10.2.15. *An n-partite pure state on* $\mathbb{H}$ *is separable if and only if Condition* 10.10 *holds.*

Proof. Taking into account Theorem 10.2.13 and Lemma 10.2.14, we need only to prove that the n-partite pure state is Schmidt decomposable and has Schmidt number 1 if and only if the rank of matrices $M_\psi^{(k)}$ are equal to 1 for $k = 1, 2, \ldots, n$. If the n-partite pure state ψ is Schmidt decomposable and has Schmidt number 1, then we have $\psi = |e_1^{(1)}\rangle \otimes |e_1^{(2)}\rangle \otimes |e_1^{(n)}\rangle$. We can calculate the density operator

$$\rho_\psi = |\psi\rangle\langle\psi| = \left|e_1^{(1)} \otimes e_1^{(2)} \otimes \cdots \otimes e_1^{(n)}\right\rangle\left\langle e_1^{(1)} \otimes e_1^{(2)} \otimes \cdots \otimes e_1^{(n)}\right|$$

and the reduced density operators

$$\begin{aligned}\rho_\psi^{(k)} &= \operatorname{tr}_{\mathbb{H}^{(k)}}(\rho_\psi)\\ &= \left|e_1^{(1)} \otimes e_1^{(2)} \otimes e_1^{(k-1)} \otimes e_1^{(k+1)} \otimes e_1^{(n)}\right\rangle\\ &\quad \left\langle e_1^{(1)} \otimes e_1^{(2)} \otimes e_1^{(k-1)} \otimes e_1^{(k+1)} \otimes e_1^{(n)}\right| \quad (k = 1, 2, \ldots, n).\end{aligned}$$

Therefore, we have

$$M_\psi^{(k)} = \begin{pmatrix} 1 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad k = 1, 2, \ldots, n.$$

This means that $M_\psi^{(k)}$ $(k = 1, 2, \ldots, n)$ only have 1 to be their nonzero eigenvalues. Therefore, Condition 10.10 holds.

On the other hand, all $M_\psi^{(k)}$ $(k = 1, 2, \ldots, n)$ only have 1 to be their nonzero eigenvalues. Let $f(k)_1$ be the eigenvector corresponding to the eigenvalues 1 for $k = 1, 2, \ldots, n$ respectively. Then we have $\psi = f_1^{(1)} \otimes f_1^{(2)} \otimes f_1^{(n)}$, which is the Schmidt decomposition of ψ with Schmidt number 1. Therefore, ψ is separable. This proves the theorem. □

The above theorem can be used to determine if an n-partite pure state is separable or entangled.

Example 10.1. Consider the n-cat state (in honor of Schrodinger's cat) $\psi = \frac{\sqrt{2}}{2}(|0^{\otimes n}\rangle + |1^{\otimes n}\rangle)$ with the Eiinsten–Podolsky–Rosen–Bohm pair $\frac{\sqrt{2}}{2}(|00\rangle + |11\rangle)$ for $n = 2$ and Greenberger–Hone–Zeilinger–Mermin state $\frac{\sqrt{2}}{2}(|000\rangle + |111\rangle)$ when $n = 3$. We have the density operator for $n = 1, 2, \ldots,$

$$\rho_\psi = \frac{1}{2}(|0^{\otimes n}\rangle\langle 0^{\otimes n}| + |0^{\otimes n}\rangle\langle 1^{\otimes n}| + |1^{\otimes n}\rangle\langle 0^{\otimes n}| + |1^{\otimes n}\rangle\langle 1^{\otimes n}|)$$

and reduced density operators

$$\rho_\psi^{(k)} = \frac{1}{2}(|0^{\otimes(n-1)}\rangle\langle 0^{\otimes(n-1)}| + |1^{\otimes(n-1)}\rangle\langle 1^{\otimes(n-1)}|).$$

This means that

$$M_\psi^{(k)} = \begin{pmatrix} \frac{1}{2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \frac{1}{2} \end{pmatrix}, \quad k = 1, 2, \ldots, n.$$

Then we have $\det(M_\psi^{(k)} - E_k) = \frac{1}{4} \neq 0$ for $k = 1, 2, \ldots, n$. Therefore, the n-cat state is an entangled (i. e., not separable) state according to Theorem 10.2.15.

In the following, we will express separability of the n-partite pure state ψ in terms of its matrix of amplitudes $M = w[a_{i_1 i_2 \cdots i_n}]$ and its reduced matrices M_k, which takes the following form for $k = 1, 2, \ldots, n$:

$$M_k = \begin{pmatrix} a_{11\cdots111\cdots111} & a_{11\cdots121\cdots111} & \cdots & a_{11\cdots1d_k1\cdots111} \\ a_{11\cdots111\cdots112} & a_{11\cdots121\cdots112} & \cdots & a_{11\cdots1d_k1\cdots112} \\ \vdots & \vdots & \vdots & \vdots \\ a_{11\cdots111\cdots11d_n} & a_{11\cdots121\cdots11d_n} & \cdots & a_{11\cdots1d_k1\cdots11d_n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i_1i_2\cdots i_{k-1}1i_{k+1}\cdots i_n} & a_{i_1i_2\cdots i_{k-1}2i_{k+1}\cdots i_n} & \cdots & a_{i_1i_2\cdots i_{k-1}d_kI_{k+1}\cdots i_n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{d_1d_2\cdots d_{k-1}1d_{k+1}\cdots d_n} & a_{d_1d_2\cdots d_{k-1}2d_{k+1}\cdots d_n} & \cdots & a_{d_1d_2\cdots d_{k-1}d_kd_{k+1}\cdots d_n} \end{pmatrix}. \tag{10.11}$$

The following lemma follows from a straightforward calculation and is therefore omitted.

Lemma 10.2.16. *The reduced matrices $M_\psi^{(k)}$, $k = 1, 2, \ldots, n$ satisfy the following relation:*

$$M_\psi^{(k)} = M_k M_k^*, \quad k = 1, 2, \ldots, n, \tag{10.12}$$

where $M_k (k = 1, 2, \ldots, n)$ are given by (10.11) *and M_k^* is the Hermitian of M_k for $k = 1, 2, \ldots, n$.*

Theorem 10.2.17. *A n-partite pure state ψ is separable if and only if the rank of matrices M_k are equal to 1 for $k = 1, 2, \ldots, n$.*

Proof. From Theorem 10.2.15 and Lemma 10.2.14, we know that a n-partite pure state ψ is separable if and only if the rank of matrices $M_\psi^{(k)}$ $(k = 1, 2, \ldots, n)$ are equal to 1. Taking into account of Lemma 10.2.16, we know that the rank of matrices $M_\psi^{(k)}$ $(k = 1, 2, \ldots, n)$ are equal to 1 if and only if the rank of M_k $(k = 1, 2, \ldots, n)$ are equal to 1. This proves the theorem. □

The following corollary follows easily from Theorem 10.2.17.

Corollary 10.2.18. *An n-partite pure state ψ is separable if and only if the determinants of all of the 2×2 submatrices of M_k $(k = 1, 2, \ldots, n)$ are zeros.*

10.3 Measurements of entanglement

Recall from Definition 10.2.5 that a quantum state ρ_{AB} on $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$ is separable if ρ_{AB} is in the convex closure of the subset of all tensor product states $\omega_A \otimes \sigma_B$, where $\omega_A \in \mathcal{S}(\mathbb{H}_A)$ and $\sigma_B \in \mathcal{S}(\mathbb{H}_B)$. A state is entangled if it is not separable.

According to Holevo–Shirokov–Werner [84, 85], a separable bipartite $\rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$ ($\mathbb{H}_{AB}$ is an infinite-dimensional Hilbert space representing the bipartite system AB) can be characterized by Lemma 3.3.7, which states that a state ρ_{AB} on $\mathbb{H}_{AB}$ is separable if and only if there exists a Borel probability measure on $\mathcal{S}(\mathbb{H}_A) \times \mathcal{S}(\mathbb{H}_B)$ such that

$$\rho_{AB} = \int\limits_{\mathcal{S}(\mathbb{H}_A)} \int\limits_{\mathcal{S}(\mathbb{H}_B)} (|\varphi\rangle_A\langle\varphi| \otimes |\psi\rangle_B\langle\psi|)\mu(d\varphi d\psi).$$

In the finite-dimensional case, application of Caratheodory's theorem reduces this to the familiar definition of a separable state as a finite convex combination of pure states. In general, we call the state ρ_{AB} *countably decomposable* if it is possible to find a representation (10.8) with a purely atomic measure μ on $\mathcal{S}(\mathbb{H}_{AB})$. In other words, if ρ_{AB} admits the following decomposition:

$$\rho_{AB} = \sum_i p_i \rho_A^{(i)} \otimes \rho_B^{(i)}, \quad p_i \geq 0, \quad \sum_i p_i = 1, \tag{10.13}$$

where $\rho_A^{(i)}$ and $\rho_B^{(i)}$ are density operators representing subsystems A and B, respectively. In the case when ρ_{AB} can be written in the above form, it is called *separable* or *classically correlated* .

Although detection of quantum entanglement is not one of the main topics to be discussed in this book, one of the most important problems we have to face is that there is no simple way of deciding if a given state ρ_{AB} is entangled or not. Other than its characterization stated in Lemma 3.3.7, the general problem of separability remains unresolved despite the fact that huge effort has been expended so far to invent stronger and easier to apply separability criteria (see, e. g., the recent review on detection of entanglement Ghne and Tuth [52]).

We need the following preliminary results for investigations of measurements of entanglement in the following three subsections. The proof is omitted.

Lemma 10.3.1 (Fannes' inequality [48]). *Let σ and ρ be states on a finite-dimensional Hilbert space $\mathbb{H}$ with $\dim(\mathbb{H}) = d < +\infty$. If $\|\sigma - \rho\|_1 < 1/3$, then*

$$\big|H(\sigma) - H(\rho)\big| \leq \log(d)\|\sigma - \rho\|_1 - \|\sigma - \rho\|_1 \log(\|\sigma - \rho\|_1). \tag{10.14}$$

Lemma 10.3.2. *Let $\mathbb{H}$ be a separable complex Hilbert space. Let ω be a state that is supported on a finite-dimensional subspace of $\mathcal{S}(\mathbb{H})$, and let $(\omega_n)_{n=1}^{+\infty}$, $\omega_n \in \mathcal{S}(\mathbb{H}^{\otimes n})$, be a sequence of states satisfying*

$$\lim_{n\to+n} \left\|\omega_n - \omega^{\otimes n}\right\|_1 = 0.$$

Then

$$H(\omega) \leq \liminf_{n\to+\infty} \frac{1}{n} H(\omega_n). \tag{10.15}$$

Proof. Let $\mathbf{P}$ be the projection operator on the support of ω, and let

$$\eta_n := \mathbf{P}^{\otimes n}\omega_n\mathbf{P}^{\otimes n}, \quad \lambda_n := \|\eta_n\|_1. \tag{10.16}$$

The trace-norm distance is nonincreasing under trace preserving with completely positive maps. Therefore,

$$\|\omega^{\otimes n} - \eta_n\|_1 + (1 - \lambda_n) \leq \|\omega^{\otimes n} - \omega_n\|_1. \tag{10.17}$$

Hence, if $\lim_{n\to+\infty} \|\omega^{\otimes n} - \omega_n\|_1 = 0$ holds, then also $\lim_{n\to+\infty} \|\omega^{\otimes n} - \eta_n\|_1 = 0$, and $\lim_{n\to+\infty} \lambda_n = 1$. In turn, if $\lim_{n\to+\infty} \|\omega^{\otimes n} - \eta_n\|_1 = 0$, then

$$\lim_{n\to+\infty} \|\lambda_n\omega^{\otimes n} - \eta_n\|_1 = 0, \tag{10.18}$$

as

$$\|\lambda_n\omega^{\otimes n} - \eta_n\|_1 \leq |\lambda_n - 1| + \|\omega^{\otimes n} - \eta_n\|_1. \tag{10.19}$$

The triangle inequality yields

$$\begin{aligned} &\frac{|H(\omega^{\otimes n}) - H(\eta_n)|}{n} \\ &\leq \frac{|H(\omega^{\otimes n}) - H(\eta_n/\lambda_n)|}{/n} + |1 - 1/\lambda_n|H(\eta_n)/n - \log_2(\lambda_n)/(n\lambda_n). \end{aligned} \tag{10.20}$$

The second term on the right-hand side in (10.20) certainly vanishes in the limit $n \to +\infty$, as $H(\eta_n)/n \leq C$ for all $n \in \mathbb{N}$ for some appropriately chosen $C > 0$. By applying

Fannes' inequality (see Lemma 10.3.1) on the first term and by making use of (10.19), one can conclude that

$$\lim_{n\to+\infty} \frac{|H(\omega^{\otimes n}) - H(\mathbf{P}^{\otimes n}\omega_n \mathbf{P}^{\otimes n})|}{n} = 0, \tag{10.21}$$

if $\lim_{n\to+\infty} \|\omega^{\otimes n} - \omega_n\|_1 = 0$. Using the function η defined as $\eta_(x) = -x\log_2(x)$, one finds that

$$\liminf_{n\to+\infty} \frac{1}{n} H(\omega_n) \geq \liminf_{n\to+\infty} \frac{1}{n} \operatorname{tr}[\mathbf{P}^{\otimes n}\eta(\omega_n)] = H(\omega),$$

which is the statement of the lemma. This proves the lemma. □

In preparation for investigation of the three measurements of entanglement (entropy of entanglement, relative entropy of entanglement and entanglement of formation) for energy-constrained quantum states, we recall the following.

Let $\mathbf{H}$ be an $\mathfrak{H}$-operator defined on the composite Hilbert space $\mathbb{H}_{AB} = \mathbb{H}_A \otimes \mathbb{H}_B$. That is, $\mathbf{H}$ is an unbounded positive linear operator on $\mathbb{H}_{AB}$ with a discrete point spectrum (eigenvalues),

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq \cdots$$

of finite multiplicity will be called an $\mathfrak{H}$-operator. The operator $\mathbf{H}$ normally denotes an Hamiltonian that represents an energy of the system. In this case, the quantity $\operatorname{tr}[\sigma\mathbf{H}]$ represents the expected energy under the state σ. For a given number $M > 0$, let $\mathcal{K}_{\mathbf{H}}(M) \subset \mathcal{S}(\mathbb{H}_{AB})$ be the set of states with a mean energy of at most M, where

$$\mathcal{K}_{\mathbf{H}}(M) := \{\sigma \in \mathcal{S}(\mathbb{H}_{AB}) \mid \operatorname{tr}[\sigma\mathbf{H}] < M\}.$$

It has been shown in Theorem 3.2.5 that $\mathcal{K}_{\mathbf{H}}(M)$ is a compact subset of $\mathcal{S}(\mathbb{H}_{AB})$ under the trace-norm $\|\cdot\|_1$.

In the following, results on measurement of entanglement of energy-constrained states in $\mathcal{K}_{\mathbf{H}}(M)$ will be explored. In particular, we will see, unlike in the unconstrained case illustrated in Proposition 10.3.4, that in sequence of energy-constrained states $(\sigma_k)_{k=1}^{+\infty} \subset \mathcal{K}_{\mathbf{H}}(M)$ that converges in trace-norm $\|\cdot\|_1$ to some state $\sigma \in \mathcal{K}_{\mathbf{H}}(M)$, the sequence of the entropies of entanglement of σ_k necessarily converge to that of entanglement of σ.

Let $\mathbf{H}$ be an $\mathfrak{H}$-operator on a separable complex Hilbert space $\mathbb{H}$. To explore asymptotic results of entropy of entanglement, relative entropy of entanglement and entanglement of formation for energy constrained state $\omega \in \mathcal{K}_{\mathbf{H}}(M)$ is obtained below. For each $n \in \mathbb{N}$, we define an $\mathfrak{H}^{(n)}$-operator on $\mathbb{H}^{\otimes n}$ as follows:

$$\mathbf{H}^{(n)} = \underbrace{\mathbf{H} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I}}_{n \text{ factors}} + \cdots + \underbrace{\mathbf{I} \otimes \mathbf{I} \otimes \cdots \otimes \mathbf{I} \otimes \mathbf{H}}_{n \text{ factors}}. \tag{10.22}$$

In the following subsections, three different types of entanglement measures for bipartite quantum states will be explored, namely, (i) entropy of entanglement; (ii) relative entropy of entanglement and (iii) entanglement of formation will be explored.

10.3.1 Entropy of entanglement

Entanglement measures give an account of the degree of entanglement of a given entangled bipartite quantum state $\rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$. One of such measures is the entropy of entanglement denoted by $\mathrm{EoE}(\cdot) : \mathcal{S}(\mathbb{H}_{AB}) \to [0, +\infty[$ is defined below.

Definition 10.3.3. For a pure state $\rho_{AB} \in \mathcal{S}(\mathbb{H}_{AB})$ of a composite system AB, its entropy of entanglement $\mathrm{EoE}(\rho_{AB})$ is given by

$$\mathrm{EoE}(\rho_{AB}) = H(\mathrm{tr}_B[\rho_{AB}])(= H(\mathrm{tr}_A[\rho_{AB}])), \tag{10.23}$$

where $H : \mathcal{S}(\mathbb{H}_A) \to [0, +\infty[$ $(H : \mathcal{S}(\mathbb{H}_B) \to [0, +\infty[)$ denotes the von Neumann entropy defined by $H(\sigma) = \mathrm{tr}[\eta(\sigma)]$,

$$\eta(x) = \begin{cases} -x \log x & \text{for } x \neq 0, \\ 0 & \text{for } x = 0. \end{cases}$$

for $\sigma \in \mathcal{S}(\mathbb{H}_A)$, and $\mathrm{tr}_A[\cdot]$ and $\mathrm{tr}_B[\cdot]$ are the partial traces of $[\cdot]$ taken over the component Hilbert space $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively.

The entropy of entanglement $\mathrm{EoE}(\cdot)$ defined above is a functional mapping composite state space $\mathcal{S}(\mathbb{H}_{AB})$ into the set of nonnegative real numbers $[0, +\infty[$; the larger the number is, the more entangled is the quantum state. For example, if $\rho_{AB} = \rho_A \otimes \rho_B$ is a product pure state, then $\mathrm{EoE}(\rho_{AB}) = H(\mathrm{tr}_B[\rho_{AB}]) = H(\rho_A) = 0$, since ρ_A is a pure state on $\mathbb{H}_A$.

The entropy of entanglement quantifies to what extent the state departs from a product state—it may be very different from zero for states and yet they can be very 'close' to pure product states in the trace-class norm as shown in the following example.

Example 10.2. Let $\sigma_0 = |\psi_0\rangle_{AB}\langle\psi_0|$, where $\psi_0 = \phi_A^0 \otimes \phi_B^0$ is the ground state of the bipartite system AB. Let $(\sigma_k)_{k=1}^{+\infty}$ be a sequence of pure states $\sigma_k = |\psi_k\rangle_{AB}\langle\psi_k|$, where

$$\psi_k = \sqrt{1 - \delta_k}\psi_0 + \sqrt{\frac{\delta_k}{k}} \sum_{i=1}^{k} \phi_A^{(i)} \otimes \phi_B^{(i)}, \tag{10.24}$$

and $\delta_k = 1/\log(k)$. In fact, the sequence $(\sigma_k)_{k=1}^{+\infty}$ converges to σ_0 in trace-norm $\|\cdot\|_1$, i. e., $\lim_{k\to+\infty} \|\sigma_k - \sigma_0\|_1 = 0$. However, $\lim_{k\to+\infty} \mathrm{EoE}(\sigma_k) = 1$, whereas $\mathrm{EoE}(\sigma_0) = 0$, since σ_0 it is a product state.

A. Unconstrained states

The following results in the remaining of this subsection are due originally to Eiser–Simon–Plenio [45].

Proposition 10.3.4. *For all $\psi \in \mathbb{H}_{AB}$ and all $\epsilon > 0$, there exist a vector $\phi \in \mathbb{H}_{AB}$ such that $\||\psi\rangle_{AB}\langle\psi| - |\phi\rangle_{AB}\langle\phi|\|_1 < \epsilon$ and yet*

$$\mathrm{EoE}(\phi) := H(\mathrm{tr}_B[|\phi\rangle_{AB}\langle\phi|]) = +\infty.$$

Proof. According to the Schmidt decomposition (by Proposition 10.1.4, which can be applied in this infinite-dimensional setting), there exists an orthonormal basis $\{\psi_A^{(n)}\}_{n=1}^{+\infty}$ of $\mathbb{H}_A$ and an orthonormal basis $\{\psi_B^{(n)}\}_{n=1}^{+\infty}$ in $\mathbb{H}_B$ such that $\psi \in \mathbb{H}_{AB}$ can be written in the form

$$\psi = \sum_{n=1}^{+\infty} \sqrt{p^{(n)}}\psi_A^{(n)} \otimes \psi_B^{(n)}, \tag{10.25}$$

where $\{p^{(n)}\}_{n=0}^{+\infty}$ forms a probability distribution. For each $n \in \mathbb{N}$, define a sequence $\{q_k^{(n)}\}_{k=1}^{+\infty}$ through $q_k^{(1)} = p_k$, and

$$q_k^{(n)} = \frac{p^{(n)} + 1/(kn\log_2(n)^2)}{\delta_k}, \quad n = 2, 3, \ldots, \tag{10.26}$$

where $\delta_k > 0$ is chosen in such a way that $\{q_k^{(n)}\}_{n=1}^{+\infty}$ is also a probability distribution. This gives rise to a sequence of state vectors $(\psi_k)_{k=1}^{+\infty} \subset \mathbb{H}_{AB}$ defined as

$$\psi_k = \sum_{n=1}^{+\infty} \sqrt{q_k^{(n)}}\psi_A^{(n)} \otimes \psi_B^{(n)} \tag{10.27}$$

The sequence $(p^{(n)})_{n=0}^{+\infty}$ is convergent by using the fact that $f : [0, 1] \to [0, +\infty[$ defined as $f(x) = -x\log_2 x$ is monotonously increasing on $[0, \epsilon]$ for some $\epsilon > 0$. One can show that $H(\mathrm{tr}_B[|\psi_k\rangle_{AB}\langle\psi_k|]) = +\infty$ for all $k \in \mathbb{N}$. However,

$$\lim_{k\to+\infty} \||\psi\rangle_{AB}\langle\psi| - |\psi_k\rangle_{AB}\langle\psi_k|\|_1 = \lim_{k\to+\infty} \sum_{n=1}^{+\infty} |p^{(n)} - q_k^{(n)}| = 0. \tag{10.28}$$

This proves the proposition. □

B. Energy-constrained states

Let $\mathbf{H}$ be an $\mathfrak{H}$-operator defined on the composite Hilbert space $\mathbb{H}_{AB} = \mathbb{H}_A \otimes \mathbb{H}_B$ and

$$\mathcal{K}_{\mathbf{H}}(M) := \{\sigma \in \mathcal{S}(\mathbb{H}_A) \mid \mathrm{tr}[\sigma\mathbf{H}] < M\}.$$

In the following, results on the entropy of entanglement of states in $\mathcal{K}_{\mathbf{H}}(M)$ will be explored. In particular, we will see, unlike in the unconstrained case illustrated in Proposition 10.3.4, that in sequence of energy-constrained states $(\sigma_k)_{k=1}^{+\infty} \subset \mathcal{K}_{\mathbf{H}}(M)$ that converges in trace-norm $\|\cdot\|_1$ to some state $\sigma \in \mathcal{K}_{\mathbf{H}}(M)$, the sequence of the entropies of entanglement of σ_k necessarily converge to that of entanglement of σ.

Contrary to the unconstrained bipartite state case (see Proposition 10.3.4), we have the following result for the EoE of constrained bipartite states.

Proposition 10.3.5. *Let* $M > 0$, *let* $\sigma \in \mathcal{K}_{\mathbf{H}}(M)$ *and let* $(\sigma_k)_{k=1}^{+\infty}$ *be a sequence of states* $\sigma_k \in \mathcal{K}_{\mathbf{H}}(M)$ *satisfying* $\lim_{k\to+\infty} \|\sigma_k - \sigma\|_1 = 0$. *Then*

$$\lim_{k\to+\infty} \left|\mathrm{EoE}(\sigma) - \mathrm{EoE}(\sigma_k)\right| = 0. \tag{10.29}$$

That is, the entropy of entanglement $\mathrm{EoE}(\cdot) : \mathcal{K}_{\mathbf{H}}(M) \to [0, +\infty[$ *is a trace-norm continuous functional.*

Proof. This statement is a consequence of a statement concerning the continuity of the von-Neumann entropy under an appropriate constraint of the energy (see Proposition 7.3.7): If $(\omega_k)_{k=1}^{+\infty}$ is a sequence of states taken from $\mathcal{K}_{\mathbf{H}}(M)$ satisfying $\omega_k \to \omega$ in trace norm for some state $\omega \in \mathcal{K}_{\mathbf{H}}(M)$, together with the above assumptions concerning the spectrum of $\mathbf{H}$, then

$$\lim_{k\to+\infty} H(\omega_k) = H(\omega). \tag{10.30}$$

The states $\omega_k := \mathrm{tr}_B[\sigma_k]$ with $\mathcal{K} := \mathbb{H}_A$ form such a sequence, since

$$\left\|\mathrm{tr}_B[\sigma_k] - \mathrm{tr}_B[\sigma]\right\|_1 \leq \|\sigma_k - \sigma\|_1, \quad \forall k \in \mathbb{N}, \tag{10.31}$$

by the contraction property of the partial trace under trace norm, and since

$$\mathrm{tr}\big[\mathbf{P}_A \mathbf{H}\, \mathrm{tr}_B[\sigma_k]\big] \leq \mathrm{tr}[\mathbf{H}\sigma_k] \leq M,$$

where $\mathbf{P}_A : \mathbb{H}_{AB} \to \mathbb{H}_A$ is a projection of $\mathbb{H}_{AB}$ onto $\mathbb{H}_A$. Hence,

$$\lim_{k\to+\infty} \left|\big(H(\mathrm{tr}_B[\sigma_k])\big) - H\big(\mathrm{tr}_B[\sigma]\big)\right| = 0, \quad \text{if } \lim_{k\to+\infty} \|\sigma_k - \sigma\|_1 = 0.$$

This proves the proposition. □

The proof of the next proposition follows exactly as that of Proposition 10.3.10 with a few small minor but necessary modifications and is therefore omitted.

Proposition 10.3.6. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on* $\mathbb{H}$ *and* $M > 0$. *Let* $\sigma \in \mathcal{K}_{\mathbf{H}}(M)$ *be a pure state that is supported on a finite-dimensional subspace of* $\mathcal{S}(\mathbb{H})$, *and let* $(\sigma_n)_{n=1}^{+\infty} \subset \mathcal{K}_{\mathbf{H}^{(n)}}(nM)$ *be a sequence of states satisfying*

$$\lim_{n\to+\infty} \|\sigma_n - \sigma\|_1 = 0. \tag{10.32}$$

Then

$$\lim_{n\to+\infty} \frac{|\mathrm{EoE}(\sigma_n) - \mathrm{EoE}(\sigma^{\otimes n})|}{n} = 0. \tag{10.33}$$

10.3.2 Relative entropy of entanglement

The following concept and properties of the set of separable states in a bipartite system AB was first introduced and investigated by Werner [176], Clifton et al. [25, 26, 59], Brocker and Werner [16], Horodecki and Lewenstein [88] and Horodeski–Cirac–Lewenstein [87].

Definition 10.3.7. The set of separable states, denoted by $\mathcal{S}_{\mathrm{sep}}(\mathbb{H}_{AB})$, is defined as the set of states $\omega \in \mathcal{S}(\mathbb{H}_{AB})$ for which there exists a sequence $(\omega_k)_{k=1}^{+\infty} \subset \mathcal{S}(\mathbb{H}_A) \otimes \mathcal{S}(\mathbb{H}_B)$, such that $\lim_{k\to+\infty} \|\omega_k - \omega\|_1 = 0$ and such that each ω_k is of the form

$$\omega_k = \sum_{i=1}^{+\infty} p_i^{(k)} \eta_A^{(k,i)} \otimes \eta_B^{(k,i)}, \tag{10.34}$$

where $\eta_A^{(k,i)} \in \mathcal{S}(\mathbb{H}_A)$, $\eta_B^{(k,i)} \in \mathcal{S}(\mathbb{H}_B)$ and $\{p_i^{(k)}\}_{i=1}^{+\infty}$ form a probability distribution for all $k \in \mathbb{N}$.

For both finite and infinite-dimensional cases, the set of separable states, $\mathcal{S}_{\mathrm{sep}}(\mathbb{H}_{AB})$, is the closed convex hull of the set of product states (with respect to the topology induced by the trace norm).

Definition 10.3.8 (Relative entropy of entanglement). The relative entropy of entanglement is the map $\mathrm{REoE} : \mathcal{S}(\mathbb{H}_{AB}) \to [0, +\infty[$ defined by

$$\mathrm{REoE}(\omega) = \inf_{\rho\in\mathcal{S}_{\mathrm{sep}}(\mathbb{H}_{AB})} H(\omega\|\rho), \tag{10.35}$$

where $H(\cdot\|\cdot) : \mathcal{S}(\mathbb{H}_{AB}) \times \mathcal{S}(\mathbb{H}_{AB}) \to [-\infty, +\infty]$ is the relative entropy defined by Definition 8.1.3.

Similar to the EoE of bipartite states for energy-constrained case (see Proposition 10.3.5), we have the following continuity result for REoE.

Proposition 10.3.9. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator and* $M > 0$. *The relative entropy of entanglement* $\mathrm{REoE} : \mathcal{K}_{\mathbf{H}}(M) \subset \mathcal{S}(\mathbb{H}_{AB}) \to [0, +\infty[$ *is continuous in trace-norm* $\|\cdot\|_1$.

Proof. Let $(\sigma_k)_{k=1}^{+\infty}$ be a sequence of states with the constraint $\mathcal{K}_{\mathbf{H}}(M)$ for which $\lim_{k\to+\infty} \|\sigma_k - \sigma\|_1 = 0$. Since $\mathcal{K}_{\mathbf{H}}(M)$ is a compact subset of $\mathcal{S}(\mathbb{H}_{AB})$. Let $\rho_k \in \mathcal{K}_{\mathbf{H}}(M)$ for each k be the state for which

$$\mathrm{REoE}(\sigma_k) = \inf_{\rho \in \mathcal{S}_{\mathrm{sep}}(\mathbb{H}_{AB})} H(\sigma_k \| \rho) = H(\sigma_k \| \rho_k)$$

(such a state exists, due to the compactness of the set $\mathcal{K}_{\mathbf{H}}(M)$ and the lower semicontinuity of the relative entropy $H(\cdot\|\cdot)$). Then

$$\big|\mathrm{REoE}(\sigma) - \mathrm{REoE}(\sigma_k)\big| \leq \big|H(\sigma) - H(\sigma_k)\big| + \big|-\mathrm{tr}[\sigma \log(\rho)] + \mathrm{tr}[\sigma_k \log(\rho_k)]\big|. \tag{10.36}$$

As $\sigma, \sigma_k \in \mathcal{K}_{\mathbf{H}}(M)$, $\lim_{k\to+\infty} |H(\sigma_k) - H(\sigma)| = 0$. The second term on the right-hand side of (10.36) is bounded from above by

$$\begin{aligned}
&\big|-\mathrm{tr}[\sigma \log(\rho)] + \mathrm{tr}[\sigma_k \log(\rho_k)]\big| \\
&\quad \leq \big|-\mathrm{tr}[\sigma \log(\rho)] + \mathrm{tr}[\sigma \log(\omega_k)]\big| \\
&\qquad + \big|-\mathrm{tr}[\sigma \log(\omega_k)] + \mathrm{tr}[\sigma_k \log(\omega_k)]\big| \\
&\qquad + \big|-\mathrm{tr}[\sigma_k \log(\omega_k)] + \mathrm{tr}[\sigma_k \log(\omega_k)]\big|,
\end{aligned} \tag{10.37}$$

where $\omega_k := \|\sigma - \sigma_k\|_1 \sigma_k + (1 - \|\sigma - \sigma_k\|_1)\rho$. Due to the operator monotonicity of the logarithm,

$$-\mathrm{tr}[\sigma \log(\rho)] + \mathrm{tr}[\sigma \log(\omega_k)] \geq \log(1 - \|\sigma - \sigma_k\|_1) \tag{10.38}$$

holds. But $-\mathrm{tr}[\sigma \log(\rho)] \leq -\mathrm{tr}[\sigma \log(\omega_k)]$ and, therefore,

$$\lim_{k\to+\infty} \big|-\mathrm{tr}[\sigma \log(\rho)] + \mathrm{tr}[\sigma \log(\omega_k)]\big| = 0. \tag{10.39}$$

In the same way, one finds that $\lim_{k\to+\infty} |-\mathrm{tr}[\sigma_k \log(\omega_k)] + \mathrm{tr}[\sigma_k \log(\rho_k)]| = 0$. The third term on the right-hand side of (10.37) can be dealt with just as in Donald and Horodeski [41], where the Gibbs state plays the role of the maximally mixed state: Since

$$\big|-\mathrm{tr}[\sigma \log(\omega - k)] + \mathrm{tr}[\sigma_k \log(\omega_k)]\big| \leq \|\sigma - \sigma_k\|_1 \big\|\log(\omega_k)\big\|_\infty, \tag{10.40}$$

one can again make use of the operator monotonicity of the logarithm to find

$$\big\|\log(\omega_k)\big\|_\infty \leq -\log(\|\sigma - \sigma_k\|_1) + \big\|\log(\sigma_\beta)\big\|_\infty, \tag{10.41}$$

and hence, $\lim_{k\to+\infty} |-\mathrm{tr}[\sigma \log(\omega_k)] + \mathrm{tr}[\sigma_k \log(\omega_k)]| = 0$. Collecting the partial results, one finds that $\lim_{k\to+\infty} |\mathrm{REoE}(\sigma) - \mathrm{REoE}(\sigma_k)| = 0$. This proves the proposition. □

The asymptotic behavior of energy-constrained REoE is explored in the following proposition.

Proposition 10.3.10. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on* $\mathbb{H}$ *and* $M > 0$. *Let* $\sigma \in \mathcal{K}_{\mathbf{H}}(M)$ *be a pure state that is supported on a finite-dimensional subspace of* $\mathcal{S}(\mathbb{H})$, *and let* $(\sigma_n)_{n=1}^{+\infty}$, $\sigma_n \in \mathcal{K}_{\mathbf{H}^{(n)}}(nM)$, *be a sequence of states satisfying* $\lim_{n\to+\infty} \|\sigma_n - \sigma^{\otimes n}\|_1 = 0$. *Then*

$$\lim_{n\to+\infty}\frac{|\mathrm{REoE}(\sigma^{\otimes n})-\mathrm{REoE}(\sigma_n)|}{n}=0. \tag{10.42}$$

Proof. The first step of the proof can be performed just as in Lemma 10.3.2. Let $\zeta \in \mathcal{K}_{\mathbf{H}}(M)$, let ω be a state that is supported on a finite-dimensional subspace of $\mathcal{K}_{\mathbf{H}}(M)$ and let $(\omega_n)_{n=1}^{+\infty}$, $\omega_n \in \mathcal{K}_{\mathbf{H}^{(n)}}(nM)$ be a sequence of states satisfying $\lim_{n\to+\infty}\|\omega_n - \omega^{\otimes n}\|_1 = 0$. Then

$$H(\zeta\|\omega)\le\liminf_{n\to+\infty}\frac{1}{n}H(\zeta^{\otimes n}\|\omega_n) \tag{10.43}$$

holds. The validity of (10.43) can be seen as follows: the relative entropy of $\zeta^{\otimes n}$ with respect to $\omega^{\otimes n}$ can be written as

$$H(\zeta^{\otimes n}\|\omega^{\otimes n})=\sup_{p\in[0,1]}\sup_{\mathbf{P}_n}\mathrm{tr}[\mathbf{P}_n(\eta(p\eta^{\otimes n}+(1-p)\omega^{\otimes n}] \tag{10.44}$$

where the second supremum of the above equation is taken over all finite-rank projection operators $\mathbf{P}_n$ on $\mathbb{H}$. So just as in Lemma 10.3.2, applying the triangle inequality and Fannes' inequality several times yields (10.42). In particular, one has to make use of the inequality

$$\begin{aligned}&\frac{|H(p\zeta^{\otimes n}+(1-p)\omega^{\otimes n})-H(p\zeta^{\otimes n}+(1-p)\eta_n)|}{n}\\ &\quad\le\frac{|H(p\zeta^{\otimes n}+(1-p)\omega^{\otimes n})-H(p\zeta^{\otimes n}+(1-p)\eta_n/\lambda_n)|}{n}\\ &\qquad+\frac{|H(p\zeta^{\otimes n}+(1-p)\eta_n/\lambda_n)-H(p\zeta^{\otimes n}+(1-p)\eta_n)|}{n}\end{aligned}$$

Equation (10.22) is now the starting point of an argument along the line of the argument of Wehrl [174] and [175]. Since the free energy can be expressed in terms of the relative entropy according to

$$\frac{1}{n}F(\omega_n,\beta,\mathbb{H}^{\otimes n}):=H(\sigma^{\otimes n}\|\omega_n)/n-\log(\mathrm{tr}[e^{-\beta\mathbf{H}}])/\beta,$$

it has the property

$$F(\omega,\beta,\mathbf{H})\le\liminf_{n\to+\infty}\frac{1}{n}F(\omega_n,\beta,\mathbf{H}^{(n)}), \tag{10.45}$$

implying that

$$\beta\,\mathrm{tr}[\omega\mathbf{H}]-H(\omega)\le\liminf_{n\to+\infty}\frac{1}{n}(\beta\,\mathrm{tr}[\omega_n\mathbf{H}^{(n)}]-H(\omega_n)) \tag{10.46}$$

and, therefore,

$$-H(\omega)\le\liminf_{n\to+\infty}(-H(\omega_n)/n)$$

$$+\beta \limsup_{n\to+\infty} \left|\mathrm{tr}[\omega_n \mathbf{H}^{(n)}]\right|/n + \beta\,\mathrm{tr}[\omega \mathbf{H}] \tag{10.47}$$

for all $\beta > 0$. Again, the sum of the last two terms of (10.47) is bounded from above by $2\beta M$, providing the inequality

$$-H(\omega) \le \liminf_{n\to+\infty}(-H(\omega_n)/n). \tag{10.48}$$

Statement (10.48) and the statement of Lemma 10.3.2 then imply that

$$H(\omega) = \lim_{n\to+\infty} H(\omega_n).$$

Finally, one can argue as in Proposition 10.3.10, by taking $\omega := \mathrm{tr}_B[\sigma]$ and $\omega_n := \mathrm{tr}_B[\sigma_n]$. This gives rise to the asymptotic continuity property stated in (10.42). This proves the proposition. □

10.3.3 Entanglement of formation

We are concerned with the entanglement of formation (EoF), which was originally introduced by Bennet et al. [7] in measuring the entanglement of a mixed state in a finite-dimensional bipartite system. In this section, we generalize the definition of EoF to infinite-dimensional composite system and prove that a general quantum state in a bipartite system is separable if and only if its entanglement of formation is zero.

The presentation of this subsection is largely based on results obtained in [7], Majewski [113] and Eisert–Simon–Plenio [45].

The original definition (see [7]) of entanglement of the formation of a mixed state is defined as the minimum average entanglement of an ensemble of pure states that represents the given mixed state. This definition can be mathematically formulated as follows: The *entanglement of formation* or EoF for a bipartite state ρ_{AB} is defined as

$$\mathrm{EoF}(\rho_{AB}) = \min_{\bar{\rho}=\rho_{AB}} \sum_i p_i H(\mathrm{tr}_A[\rho_i]) \tag{10.49}$$

where the minimization is over all discrete ensembles $\mu = \{p_i, \rho_i\} \in \mathcal{P}(\mathcal{S}(\mathbb{H}_{AB}))$ having probability distributions p_i on rank-one density operators ρ_i such that its barycenter (or average state) $\bar{\rho}(\mu) := \sum_i p_i \rho_i = \rho_{AB}$.

The alternative approach to the definition of the EoF is considered in Majewski [113] in the case of tensor product of two systems $\mathbb{H}_A$ and $\mathbb{H}_B$ with one of them being finite-dimensional. By using the results of Subsection 3.3, we can generalize this approach and define the EoF in the general case by the following definition.

Definition 10.3.11 (Entanglement of formation).

$$\mathrm{EoF}(\rho) = \inf_{\mu\in\mathcal{P}_{\{\rho\}}(\mathcal{S}(\mathbb{H}))} \int\limits_{\mathcal{S}(\mathbb{H})} H(\mathrm{tr}_A[\rho])\mu(d\rho), \tag{10.50}$$

where $\mathcal{P}_{\{\rho\}}(\mathcal{S}(\mathbb{H}))$ is the collection of $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$ with barycenter $\bar{\rho}(\mu) = \rho$.

A. Unconstrained states

For unconstrained states, we state the following main result of EoF without a proof.

Theorem 10.3.12. *The state $\omega \in \mathcal{S}(\mathbb{H}_{AB})$ is separable if and only if* $\mathrm{EoF}(\omega) = 0$.

B. Energy-constrained states

For a given mixed state $\sigma \in \mathcal{K}_{\mathbf{H}}(M)$, $M > 0$, there exist (countably infinitely many) sequences $(p^{(i)})_{i=1}^{+\infty}$ of positive numbers forming a probability distribution, $p^{(1)} \geq p^{(2)} \geq \cdots$, and sequences $(\psi^{(i)})_{i=1}^{+\infty}$ of state vectors $\psi^{(i)} \in \mathbb{H}$ such that

$$\sigma = \sum_i p^{(i)} |\psi^{(i)}\rangle_{\mathbb{H}} \langle \psi^{(i)}|. \tag{10.51}$$

The pair $(\{p^{(i)}\}_{i=1}^{+\infty}, \{\psi^{(i)}\}_{i=1}^{+\infty})$ will be called decomposition of σ. As for a finite-dimensional Hilbert space, one may define the entanglement of formation as

$$\mathrm{EoF}(\sigma) = \inf \sum_i p_i H(\mathrm{tr}_B[|\psi^{(i)}\rangle_{\mathbb{H}} \langle \psi^{(i)}|]), \tag{10.52}$$

where the infimum is understood to be with respect to all decompositions of σ. In the case of a finite-dimensional Hilbert space, the infimum is always attained, and by virtue of Caratheodory's theorem, one can find an upper bound for the required number of terms in a decomposition of the state. It is worth noting that, using a different language, a decomposition of a state σ can also be represented by probability measures μ_σ on state space with the barycenter $\bar{\rho}(\mu_\sigma) = \int_{\mathcal{S}(\mathbb{H}_{AB})} \rho\mu_\sigma(d\rho)$.

A convergence result for EoF for energy-constrained bipartite state is given below.

Proposition 10.3.13. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on* $\mathbb{H}_{AB}$ *and let* $M > 0$. *Let* $\sigma = |\psi\rangle_{AB}\langle\psi| \in \mathcal{K}_{\mathbf{H}}(M)$ *be a pure state, and let* $(\sigma_k)_{k=1}^{+\infty}$ *be a sequence of states in* $\mathcal{K}_{\mathbf{H}}(M)$ *with* $\lim_{k\to+\infty} \|\sigma_k - \sigma\|_1 = 0$. *Then*

$$\lim_{k\to+\infty} \big|\mathrm{EoF}(\sigma_k) - \mathrm{EoF}(\sigma)\big| = 0. \tag{10.53}$$

Proof. We start with proving the lower semicontinuity of $\mathrm{EoF}(\cdot)$ in σ. Let $r > 0$ be a number satisfying $\mathrm{EoF}(\sigma_k) \leq r$ for all $k \in \mathbb{N}$. For all $\epsilon > 0$, there exists a decomposition $(\{p_k^{(i)}\}_{i=1}^{+\infty}, \{\psi_k^{(i)}\}_{i=1}^{+\infty})$ of each σ_k such that

$$\sum_i p_k^{(i)} H(\mathrm{tr}_B[|\psi_k^{(i)}\rangle_{AB}\langle\psi^{(i)}|]) \leq r + \epsilon. \tag{10.54}$$

The fact that $\sigma_k \to \sigma = |\psi\rangle_{AB}\langle\psi|$ in trace norm implies that there exists a sequence of real numbers $(r_k)_{k=1}^{+\infty}$ with $\lim_{k\to+\infty} r_k = 0$, and such that

$$\lim_{k\to+\infty} \sum_i p^{(i)} \theta(r_k - \||\psi_k^{(i)}\rangle_{AB}\langle\psi_k^{(i)} - \sigma\|_1) = 1, \tag{10.55}$$

where $\theta : \mathbb{R} \to \{0, 1\}$ is the Heaviside function. For each k, construct a sequence $(q_k^{(i)})_{i=1}^{+\infty}$ of real numbers as

$$q_k^{(i)} = \begin{cases} p_k^{(i)}, & \text{if } r_k - \||\psi_k^{(i)}\rangle_{AB}\langle\psi_k^{(i)}| - \sigma\|_1 > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{10.56}$$

Similarly, define a sequence $(\phi^{(i)})_{i=1}^{+\infty}$ through

$$\phi_k^{(i)} = \begin{cases} \psi_k^{(i)}, & \text{if } r_k - \||\psi_k^{(i)}\rangle_{AB}\langle\psi_k^{(i)}| - \sigma\|_1 > 0, \\ \psi, & \text{otherwise.} \end{cases}$$

It follows that

$$\sum_i q_k^{(i)} H(\mathrm{tr}_B[|\psi_k^{(i)}\rangle_{AB}\langle\psi_k^{(i)}|]) < r + \epsilon, \tag{10.57}$$

$\lim_{k\to+\infty} \||\psi_k^{(i)}\rangle_{AB}\langle\psi_k^{(i)}| - \sigma\|_1 = 0$ for all i, and $\lim_{k\to+\infty} \sum_i q_k^{(i)} = 1$. Hence, due to the lower semicontinuity of the von Neumann entropy we can conclude that also $H(\mathrm{tr}_B[\sigma]) \leq r + \epsilon$. As $\epsilon > 0$ was arbitrary, it follows that $\mathrm{EoF}(\sigma) \leq r$, which means that $\mathrm{EoF}(\cdot)$ is lower semicontinuous in σ. Note that the constraint on the mean energy was not needed in this step. The second part of the proof will be concerned with the upper semicontinuity in σ. Let $r \in \mathbb{R}$ such that $\mathrm{EoF}(\sigma_k) \geq r$ for all $k \in \mathbb{N}$. Essentially, the proof here follows to a large extent along the line of the argument in Wehrl [174] and [175], where additionally, we make use of convexity of the relative entropy functional. Let $(\{p_k^{(i)}\}_{i=1}^{+\infty}, \{\psi_k^{(i)}\}_{i=1}^{+\infty})$ be a decomposition of σ_k, and let $\omega := \mathrm{tr}_B[\sigma]$, $\omega_k := \mathrm{tr}_B[\sigma_k]$ and $\omega_i^{(i)} := \mathrm{tr}_B[|\psi_k^{(i)}\rangle_{AB}\langle\psi_k^{(i)}|]$. On using both the lower semicontinuity and the convexity of the relative entropy functional, one obtains

$$H(\omega\|\sigma_\beta) \leq \liminf_{k\to+\infty} H(\omega_k\|\sigma_\beta) \leq \liminf_{k\to+\infty} \sum_i p_k^{(i)} H(\omega^{(i)}\|\sigma_\beta). \tag{10.58}$$

Therefore,

$$\beta\,\mathrm{tr}[\omega\mathbf{H}_A] - H(\omega) \leq \liminf_{k\to+\infty}\Big(\beta\,\mathrm{tr}[\omega_k\mathbf{H}_A] - \sum_i k_k^{(i)} H(\omega_k^{(i)})\Big) \tag{10.59}$$

and

$$-H(\omega) \leq \liminf_{k\to+\infty} p_k^{(i)}(-H(\omega_k^{(i)})) + \beta \liminf_{k\to+\infty} \mathrm{tr}[\omega_k \mathbf{H}_A] - \beta\,\mathrm{tr}[\omega \mathbf{H}_A]. \tag{10.60}$$

Hence, we arrive at

$$-H(\omega) \leq \liminf_{k\to+\infty} \sum_i p_k^{(i)}(-H(\omega_j^{(i)})). \tag{10.61}$$

As $\mathrm{EoF}(\sigma_k) \geq r$ for all $k \in \mathbb{N}$, and the above decomposition is not necessarily optimal for $\mathrm{EoF}(\sigma_k)$,

$$\sum_i p_k^{(i)} H(\mathrm{tr}_B[|\psi_k^{(i)}\rangle_{AB}\langle\psi^{(i)}|]) \geq r. \tag{10.62}$$

The last step is to see that $H(\mathrm{tr}_B[\sigma]) \geq r$, which means that $\mathrm{EoF}(\cdot) : \mathcal{K}_{\mathbf{H}}(M) \to [0, +\infty[$ is also upper semicontinuous. This proves the proposition. □

Therefore, we conclude that the entanglement of formation for pure states can indeed by simply identified with the entropy of entanglement on the set $\mathcal{K}_{\mathbf{H}}(M)$. A similar argument applies again on the asymptotic limit: if we have a series of mixed states that converges in trace-norm to the n-fold tensor product of a pure state with a finite support, then, again, one can expect an asymptotic continuity as in Proposition 10.3.10.

Proposition 10.3.14. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on* $\mathbb{H}$ *and* $M > 0$*. Let* $\sigma \in \mathcal{K}_{\mathbf{H}}(M)$ *be a pure state that is supported on a finite-dimensional subspace of* $\mathcal{S}(\mathbb{H})$*, and let* $(\sigma_n)_{n=1}^{+\infty} \subset \mathcal{K}_{\mathbf{H}^{(n)}}(nM)$*, be a sequence of states satisfying*

$$\lim_{n\to+\infty} \|\sigma_n - \sigma^{\otimes n}\|_1 = 0.$$

Then

$$\lim_{n\to+\infty} \frac{|\mathrm{EoF}(\sigma_n) - \mathrm{EoF}(\sigma^{\otimes n})|}{n} = 0. \tag{10.63}$$

Proof. One can proceed in the same way as Proposition 10.3.13. Instead of the mere lower semicontinuity of the von Neumann entropy, one has to make use of the statement of Lemma 10.3.2. This proves the proposition. □

11 Quantum mutual and coherent information

This chapter defines and explores the properties of quantum mutual and coherent information. The classical capacities and quantum capacities of quantum channels can be expressed in terms of this information in later chapters.

Roughly speaking, quantum mutual information, or von Neumann mutual information, after John von Neumann, is a measure of correlation between subsystems of quantum state. It is the quantum mechanical analog of Shannon mutual information, whereas coherent information is an entropy measure used in quantum information theory. It is a property of a quantum state ρ and a quantum channel Φ and it attempts to describe how much of the quantum information in the state will remain after the state goes through the channel. In this sense, it is intuitively similar to the mutual information of classical information theory.

The presentation of this chapter is largely based on works by Holevo and Shirokov [82], Shirokov [150], Davies [30], Barnum–Nielsen–Schumacher [5], Bennett–Shor–Smolin–Thapliyal [8], Devetak [37], Wehrl [175] and Kuznetsova [102].

11.1 Quantum mutual information

Motivated by the fact that classical mutual information $I(X : Y)$ plays an important role in the computation of channel capacity in classical communication, we explore below the concept and properties of $I_m(\rho, \Phi)$, the quantum mutual information of the quantum channel Φ at the state ρ, in order to extend this classical result to its quantum counterpart. In quantum information theory, quantum mutual information is a measure of correlation between subsystems of quantum state. It is the quantum mechanical analog of Shannon mutual information.

In what follows, let $\mathbb{H}$ be a separable complex Hilbert space in finite or infinite dimensions. Let $\mathfrak{T}(\mathbb{H})$ be the Banach space of trace-class operators under trace-class norm $\|\cdot\|_1$, so that $\mathcal{S}(\mathbb{H}) \subset \mathfrak{T}_+(\mathbb{H})$, where $\mathfrak{T}_+(\mathbb{H})$ is the positive cone of $\mathfrak{T}(\mathbb{H})$.

Recall from Definition 7.1.1 that the von Neumann entropy of a state $\rho \in \mathcal{S}(\mathbb{H})$ is defined as $H(\rho) = \mathrm{tr}[\eta(\rho)]$, where the function $\eta : \mathbb{R}_+ \to \mathbb{R}$ is defined as

$$\eta(x) = \begin{cases} -x \log x & \text{for } x > 0 \\ 0 & \text{for } x = 0, \end{cases}$$

and as mentioned earlier log denotes the logarithmic function based 2.

As noted in Section 7.4, the von Neumann entropy $H(\cdot)$ can be extended from $\mathcal{S}(\mathbb{H})$ to $\mathfrak{T}_+(\mathbb{H})$ as

$$H(\mathbf{A}) = \mathrm{tr}[\eta(\mathbf{A})] - \eta(\mathrm{tr}[\mathbf{A}]), \quad \forall \mathbf{A} \in \mathfrak{T}_+(\mathbb{H}). \tag{11.1}$$

https://doi.org/10.1515/9783110788105-011

We note that the above definition is an extension from $\mathcal{S}(\mathbb{H})$ to $\mathfrak{T}_+(\mathbb{H})$, because if $\rho \in \mathcal{S}(\mathbb{H})$ then $\mathrm{tr}[\rho] = 1$ and $\eta(\mathrm{tr}[\rho]) = 0$ and $H(\rho) = \mathrm{tr}[\eta(\rho)] := S(\rho)$ reduces to the one given in Definition 7.1.1. In addition, the relative entropy $H(\cdot\|\cdot)$ on $\mathcal{S}(\mathbb{H})\times\mathcal{S}(\mathbb{H})$ defined in Definition 8.1.3 is extended to $\mathfrak{T}_+(\mathbb{H}) \times \mathfrak{T}_+(\mathbb{H})$ as

$$H(\mathbf{A}\|\mathbf{B}) = \begin{cases} \sum_{i=1}^{+\infty} \langle e_i|(\mathbf{A}\log\mathbf{A} - \mathbf{A}\log\mathbf{B} + \mathbf{B} - \mathbf{A})|e_i\rangle, & \mathrm{supp}(\mathbf{A}) \subseteq \mathrm{supp}(\mathbf{B}), \\ +\infty, & \text{otherwise}, \end{cases} \tag{11.2}$$

where $(e_i)_{i=1}^{+\infty}$ is an orthonormal basis of eigenvectors of the operator $\mathbf{A}$ and the series consists nonnegative terms.

Consider the quantum systems A, B and E described by Hilbert spaces $\mathbb{H}_A$, $\mathbb{H}_B$ and $\mathbb{H}_E$, respectively.

Suppose we are given a quantum state $\rho_A \in \mathcal{S}(\mathbb{H}_A)$ with its spectral decomposition

$$\rho_A = \sum_{i=1}^{+\infty} \lambda_i|e_i\rangle_A\langle e_i|, \quad \lambda_i \in \mathbb{C}, i = 1, 2, \dots,$$

where $|e_i\rangle_A\langle e_i| = |e_i\rangle_{\mathbb{H}_A}\langle e_i|$. Then ρ_{RA} defined by

$$\rho_{RA} = \sum_{i=1}^{+\infty} \sqrt{\lambda_i}|e_i\rangle_R \otimes |e_i\rangle_A$$

is a purification of ρ_A in the composite system RA (see (5.25) and Subsection 5.5) for definition and property of purification of quantum states), where $(|e_i\rangle_A)_{i=1}^{+\infty}$ and $(|e_i\rangle_R)_{i=1}^{+\infty}$ are orthonormal bases of the input system A and the reference system R, respectively.

Definition 11.1.1. Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel from system A to system B, and let ρ be an arbitrary quantum state in $\mathcal{S}(\mathbb{H}_A)$ with a spectral decomposition $\rho = \sum_{i=1}^{+\infty} \lambda_i|e_i\rangle_A\langle e_i|$. The *mutual information* of the channel Φ at the state ρ is defined in terms of relative entropy $H(\cdot\|\cdot)$ as follows:

$$I_m(\rho, \Phi) = H\big((\Phi \otimes \mathfrak{I}_R)\big(|\phi_\rho\rangle_{AR}\langle\phi_\rho|\big)\|\Phi(\rho) \otimes \rho\big) \tag{11.3}$$

where

$$\phi_\rho = \sum_{i=1}^{+\infty} \sqrt{\lambda_i}|e_i\rangle_A \otimes |e_i\rangle_R \in \mathbb{H}_{AR} := \mathbb{H}_A \otimes \mathbb{H}_R$$

is the purification vector for the state ρ and $\mathfrak{I}_R(\cdot)$ is the identity operator on $\mathcal{S}(\mathbb{H}_R)$.

Note that the above definition of $I_m(\rho, \Phi)$ depends neither on the choice of a reference system R represented by the space $\mathbb{H}_R$ nor on a purification vector ϕ_ρ, where $\mathrm{tr}_R[|\phi_\rho\rangle_{AR}\langle\phi_\rho|] = \rho$. This can be shown by using well-known relation between different purification vectors of a given state (see (5.25), Subsection 5.5, and properties of the relative entropy (Definition 8.1.3)).

We need the following result regarding purification of quantum states.

Lemma 11.1.2. *Let $\mathbb{H}_A$ be a separable complex Hilbert space. For an arbitrary sequence $(\rho_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H}_A)$ converging to a state ρ_0 under the trace class norm $\|\cdot\|_1$, there exists a corresponding purification sequence $(\hat{\rho}_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H}_{AR})$ converging to a purification $\hat{\rho}_0 \in \mathcal{S}(\mathbb{H}_{AR})$ of the state $\rho_0 \in \mathcal{S}(\mathbb{H}_A)$.*

Proof. The assertion of the lemma follows from the inequality (see Proposition 6.1.12)

$$\beta^2(\rho,\sigma) \le \|\rho - \sigma\|_1$$

for the Bures distance $\beta(\rho,\sigma) = \inf \|\varphi_\rho - \varphi_\sigma\|_1$, where the infimum is over all purification vectors φ_ρ and φ_σ of the states ρ and σ. This proves the lemma. □

Recall that $\mathfrak{QC}(A,B)$ denotes the set of all quantum channels from $\mathcal{S}(\mathbb{H}_A)$ to $\mathcal{S}(\mathbb{H}_B)$ endowed with the strong convergence topology, where a sequence $(\Phi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A,B)$ is said to converge strongly to a channel $\Phi_0 \in \mathfrak{QC}(A,B)$ if

$$\lim_{n\to+\infty} \Phi_n(\rho) = \Phi_0(\rho), \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A).$$

Lemma 11.1.3. *Assume that $\dim(\mathbb{H}_B) < +\infty$. Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel and let ρ_0 be a state in $\mathcal{S}(\mathbb{H}_A)$ with a spectral representation $\rho_0 = \sum_{i=1}^{+\infty} \lambda_i |e_i\rangle_A\langle e_i|$. Let*

$$\rho_n = \frac{1}{\mu_n}\sum_{i=1}^{n} \lambda_i |e_i\rangle_A\langle e_i|, \quad \textit{where } \mu_n = \sum_{i=1}^{n} \lambda_i,$$

for every n. Then $\lim_{n\to+\infty} I_m(\rho_n,\Phi) = I_m(\rho_0,\Phi)$.

Proof. Let $\mathbf{P}_n = \sum_{i=1}^{n} |e_i\rangle_A\langle e_i|$ for $n = 1,2,\ldots$. Then $(\mathbf{P}_n)_{n=1}^{+\infty}$ is an increasing sequence of finite rank projections that converges to the identity operator $\mathbf{I}_A$ on $\mathbb{H}_A$. Since $\dim(\mathbb{H}_B) < \infty$, the value

$$\begin{aligned} I_m^{(n)} &= H\big((\Phi \otimes \mathfrak{I}_R)(\hat{\rho}_n)\|\Phi(\rho_0)\otimes \rho_n\big) \\ &= \frac{1}{\mu_n} H\big(\mathbf{Q}_n((\Phi\otimes\mathfrak{I}_R)(\hat{\rho}_0))\mathbf{Q}_n \| \mathbf{Q}_n(\Phi(\rho_0)\otimes\rho_n)\mathbf{Q}_n\big) < +\infty, \end{aligned}$$

where

$$\begin{aligned} \hat{\rho}_0 &= \left(\sum_{i=1}^{+\infty} \sqrt{\lambda_i}|e_i\rangle_A\langle e_i|\right) \otimes \left(\sum_{j=1}^{+\infty} \sqrt{\lambda_j}|e_j\rangle_R\langle e_j|\right) \\ &= \sum_{i,j=1}^{+\infty} \sqrt{\lambda_i\lambda_j}|e_i\rangle_A\langle e_j| \otimes |e_i\rangle_R\langle e_j|, \end{aligned}$$

$$\hat{\rho}_n = \frac{1}{\mu_n}\sum_{i,j=1}^{n} \sqrt{\lambda_i\lambda_j}|e_i\rangle_A\langle e_j| \otimes |e_i\rangle_R\langle e_j|, \quad \text{and} \quad \mathbf{Q}_n = \mathbf{I}_B \otimes \mathbf{P}_n.$$

According to Lemma 8.2.5, we have

$$\lim_{n\to+\infty} I_m^{(n)} = H\big((\Phi\otimes\mathfrak{I}_R)(\hat\rho_0)\|\Phi(\rho_0)\otimes\rho_0\big) = I_m(\rho_0,\Phi). \tag{11.4}$$

We now prove that $\lim_{n\to+\infty} I_m^{(n)} = \lim_{n\to+\infty} I_m(\rho_n,\Phi)$ by considering the difference $I_m^{(n)} - I_m(\rho_n,\Phi)$. Since we have $H((\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)) < +\infty$, we have

$$\begin{aligned}
&I_m^{(n)} - I_m(\rho_n,\Phi)\\
&\quad= H\big((\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\|\Phi(\rho_0)\otimes\rho_n\big) - H\big((\Phi\otimes\mathfrak{I}_R)(\hat\rho_0)\|\Phi(\rho_n)\otimes\rho_n\big)\\
&\quad= -H\big((\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\big) - \operatorname{tr}\big[(\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\big)\log\big(\Phi(\rho_0)\otimes\rho_n\big)\big]\\
&\qquad+ H\big((\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\big) + \operatorname{tr}\big[(\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\log\big(\Phi(\rho_n)\otimes\rho_n\big)\big]\\
&\quad= \text{(i)} - \text{(ii)},
\end{aligned}$$

where

$$\begin{aligned}
\text{(i)} &= -\operatorname{tr}\big[(\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\log\big(\Phi(\rho_0)\otimes\rho_n\big)\big],\\
\text{(ii)} &= -\operatorname{tr}\big[(\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\log\big(\Phi(\rho_n)\otimes\rho_n\big)\big].
\end{aligned}$$

We use the following property of the logarithm of the tensor product of quantum states:

$$\log(\rho\otimes\sigma) = \log(\rho)\otimes\mathbf{I}_R + \mathbf{I}_A\otimes\log(\sigma), \tag{11.5}$$

where, in the case of nonfull-rank states ρ and σ, restrictions to the subspaces $\operatorname{supp}(\rho)$ and $\operatorname{supp}(\sigma)$ are considered, i. e.,

$$\begin{aligned}
&(\mathbf{P}_\rho\otimes\mathbf{P}_\sigma)\big(\log(\rho\otimes\sigma)\big)\\
&\quad= (\mathbf{P}_\rho\otimes\mathbf{P}_\sigma)\big(\log(\rho)\otimes\mathbf{I}_R + \mathbf{I}_A\otimes\log(\sigma)\big)\\
&\quad= \big(\mathbf{P}_\rho\log(\rho)\mathbf{P}_\rho\big)\otimes\mathbf{P}_\sigma + \mathbf{P}_\rho\otimes\big(\mathbf{P}_\sigma\log(\sigma)\mathbf{P}_\sigma\big),
\end{aligned} \tag{11.6}$$

where $\mathbf{P}_\rho$ and $\mathbf{P}_\sigma$ are respectively the projectors onto $\operatorname{supp}(\rho)$ and $\operatorname{supp}(\sigma)$. Since $\mathbf{P}_{\Phi(\rho_n)} \le \mathbf{P}_{\Phi(\rho_0)}$, we have

$$\begin{aligned}
\text{(i)} &= -\operatorname{tr}\big[(\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\log\big(\Phi(\rho_0)\otimes\mathfrak{I}_R\big)\big] - \operatorname{tr}\big[\big((\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\big)\big(\mathbf{I}_B\otimes\log(\rho_n)\big)\big]\\
&= -\operatorname{tr}\big[\Phi(\rho_n)\log\Phi(\rho_0)\big] + H(\rho_n),\\
\text{(ii)} &= -\operatorname{tr}\big[(\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\big)\big(\log\big(\Phi(\rho_n)\otimes\mathfrak{I}_R\big)\big] - \operatorname{tr}\big[(\Phi\otimes\mathfrak{I}_R)(\hat\rho_n)\big)\big(\mathbf{I}_B\otimes\log(\rho_n)\big)\big]\\
&= -H\big(\Phi(\rho_n)\big) + H(\rho_n).
\end{aligned}$$

Hence,

$$I_m^{(n)} - I_m(\rho_n,\Phi) = (i) - (ii) = -\operatorname{tr}\big[\Phi(\rho_n)\log\Phi(\rho_0)\big] - H\big(\Phi(\rho_n)\big) = H\big(\Phi(\rho_n)\|\Phi(\rho_0)\big).$$

By monotonicity of the relative entropy (see Lemma 9.1.3), we have

$$H(\Phi(\rho_n)\|\Phi(\rho_0)) \le H(\rho_n\|\rho_0) = -\sum_{i=1}^{n}\frac{\lambda_i}{\mu_n}\log\mu_n = -\log\mu_n.$$

Since $\mu_n \to 1$ as $n \to +\infty$, we get $\lim_{n\to+\infty}(I_m^{(n)} - I_m(\rho_n.\Phi)) = 0$. Taking into account (11.4), this implies that $\lim_{n\to+\infty} I_m(\rho_n,\Phi) = I_m(\rho_0,\Phi) \le +\infty$. This proves the lemma. □

The following proposition due originally to Holevo and Shirokov [82] holds for infinite-dimensional Hilbert spaces $\mathbb{H}_A$, $\mathbb{H}_B$, etc.

Proposition 11.1.4. *The map $(\rho,\Phi) \mapsto I_m(\rho,\Phi)$ is nonnegative and lower semicontinuous on the set $\mathcal{S}(\mathbb{H}_A)\times\mathfrak{QC}(A,B)$. It has the following properties:*
1. Concavity *in ρ: $I_m(p\rho_1+(1-p)\rho_2,\Phi) \ge pI_m(\rho_1,\Phi)+(1-p)I_m(\rho_2,\Phi)$ for all $0\le p\le 1$, $\rho_1,\rho_2 \in \mathcal{S}(\mathbb{H}_A)$ and $\Phi \in \mathfrak{QC}(A,B)$;*
2. Convexity *in Φ: $I_m(\rho,p\Phi_1+(1-p)\Phi_2) \le pI_m(\rho,\Phi_1)+(1-p)I_m(\rho,\Phi_2)$ for all $0\le p\le 1$, $\rho \in \mathcal{S}(\mathbb{H}_A)$ and $\Phi_1,\Phi_2 \in \mathfrak{QC}(A,B)$;*
3. The first chain rule*: for arbitrary channels $\Phi \in \mathfrak{QC}(A,B)$ and $\Psi \in \mathfrak{QC}(B,C)$, the inequality $I_m(\rho,\Psi\circ\Phi) \le I_m(\rho,\Phi)$ holds for any $\rho \in \mathcal{S}(\mathbb{H}_A)$;*
4. The second chain rule*: for arbitrary channels $\Phi \in \mathfrak{QC}(A,B)$ and $\Psi \in \mathfrak{QC}(B,C)$, the inequality $I_m(\rho,\Psi\circ\Phi) \le I_m(\Phi(\rho),\Psi)$ holds for any $\rho \in \mathcal{S}(\mathbb{H}_A)$;*
5. Subadditivity*: for arbitrary channels $\Phi \in \mathfrak{QC}(A,B)$ and $\Psi \in \mathfrak{QC}(C,D)$, the inequality*

$$I_m(\omega,\Phi\otimes\Psi) \le I_m(\omega_A,\Phi) + I_m(\omega_C,\Psi)$$

holds for any $\omega \in \mathcal{S}(\mathbb{H}_{AC})$, where $\omega_A = \mathrm{tr}_C[\omega]$ and $\omega_C = \mathrm{tr}_A[\omega]$.

Proof. By Definition 11.1.1, nonnegativity of the value $I_m(\rho,\Phi)$ follows from nonnegativity of the relative entropy $H(\cdot\|\cdot)$. By Lemma 11.1.2 and the fact that the mutual information is independent of choice of the reference system R and the purified vector ϕ_ρ, lower semicontinuity of the map $(\rho,\Phi) \mapsto I_m(\rho,\Phi)$ follows from lower semicontinuity of the relative entropy in both arguments.

1. To prove concavity of the function $\rho \mapsto I_m(\rho,\Phi)$ for each fixed $\Phi \in \mathfrak{QC}(A,B)$, we first assume that $\dim(\mathbb{H}_B) < +\infty$. Let $\rho = p\sigma_1+(1-p)\sigma_2$ for $0\le p\le 1$ and let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be an increasing sequence of finite-rank spectral projectors of the state ρ strongly converging to the identity operator $\mathbf{I}_A$ on the Hilbert space $\mathbb{H}_A$. Let

$$\rho_n = \frac{\mathbf{P}_n\rho\mathbf{P}_n}{\mathrm{tr}[\mathbf{P}_n\rho]} = \frac{p\mathbf{P}_n\sigma_1\mathbf{P}_n + (1-p)\mathbf{P}_n\sigma_2\mathbf{P}_n}{p\,\mathrm{tr}[\mathbf{P}_n\sigma_1] + (1-p)\,\mathrm{tr}[\mathbf{P}_n\sigma_2]} = \frac{\mu_1^n\sigma_1^n + \mu_2^n\sigma_2^n}{\mu_1^n + \mu_2^n},$$

where

$$\mu_1^n = p\,\mathrm{tr}[\mathbf{P}_n\sigma_1], \quad \sigma_1^n = p\frac{\mathbf{P}_n\sigma_1\mathbf{P}_n}{\mu_1^n},$$
$$\mu_2^n = (1-p)\,\mathrm{tr}[\mathbf{P}_n\sigma_2], \quad \sigma_2^n = (1-p)\frac{\mathbf{P}_n\sigma_2\mathbf{P}_n}{\mu_2^n}.$$

By the concavity of the map $\rho \mapsto I_m(\rho,\Phi)$ on the set $\mathcal{S}_f(\mathbb{H}_A)$ ($\mathcal{S}_f(\mathbb{H}_A)$ is the set of quantum states on $\mathbb{H}_A$ that has finite ranks), as mentioned earlier before Proposition 11.1.4, we have

$$I_m(\rho_n,\Phi) \geq \frac{\mu_1^n}{\mu_1^n+\mu_2^n}I_m(\sigma_1,\Phi) + \frac{\mu_2^n}{\mu_1^n+\mu_2^n}I_m(\sigma_2,\Phi).$$

2. Convexity of the function $\Phi \mapsto I_m(\rho,\Phi)$ follows from joint convexity of the relative entropy $H(\cdot\|\cdot)$ in its arguments.

3. The first chain rule immediately follows from Definition 11.1.1 and monotonicity of the relative entropy $H(\cdot\|\cdot)$.

4. The second chain rule is also proved by using monotonicity of the relative entropy in the following way:

Let $|\varphi_\rho\rangle_{AR}\langle\varphi_\rho|$ be a purification of the state $\rho \in \mathcal{S}(\mathbb{H}_A)$ in the space $\mathbb{H}_{AR}$; then

$$\Phi(|\varphi_\rho\rangle_{AR}\langle\varphi_\rho|) = (\mathbf{V}\otimes\mathbf{I}_R)|\varphi_\rho\rangle_{AR}\langle\varphi_\rho|(\mathbf{V}\otimes\mathbf{I}_R)$$

is a purification of the state $\Phi(\rho) \in \mathcal{S}(\mathbb{H}_B)$ in the space $\mathbb{H}_B\otimes\mathbb{H}_E\otimes\mathbb{H}_R$ (here $\mathbf{V}$ is the isometry from Stinespring representation (see Theorem 4.3.4) of the channel Φ). Hence, Lemma 11.1.3 implies that $\lim_{n\to+\infty} I_m(\rho_n,\Phi) = I_m(\rho,\Phi)$. By using the lower semicontinuity of the function $\rho \mapsto I_m(\rho,\Phi)$, we obtain

$$\begin{aligned} I_m(\rho,\Phi) &\geq \liminf_{n\to+\infty}\frac{\mu_1^n}{\mu_1^n+\mu_2^n}I_m(\sigma_1,\Phi) + \liminf_{n\to+\infty}\frac{\mu_2^n}{\mu_1^n+\mu_2^n}I_m(\sigma_2,\Phi) \\ &\geq pI_m(\sigma_1,\Phi) + (1-p)I_m(\sigma_2,\Phi). \end{aligned}$$

Let Φ be an arbitrary quantum channel. Consider the sequence of channels $(\Phi_n)_{n=1}^{+\infty}$, where $\Phi_n = \Pi_n\circ\Phi$ for each n, with a finite-dimensional output, where

$$\Pi_n(\rho) = \mathbf{P}_n\rho\mathbf{P}_n + \mathrm{tr}\big[((\mathbf{I}_B-\mathbf{P}_n)\rho)|\psi\rangle_B\langle\psi|\big]$$

is a quantum channel from $\mathcal{S}(\mathbb{H}_B)$ to itself for each n, the sequence $(\mathbf{P}_n)_{n=1}^{+\infty}$ is an increasing sequence of finite rank projections strongly converge to the identity operator $\mathbf{I}_B$ on $\mathbb{H}_B$ and $|\psi\rangle_B\langle\psi|$ is a pure state on $\mathbb{H}_B$. Then for each n the function $\rho \mapsto I_m(\rho,\Phi_n)$ is concave by the above observation. Since

$$I_m(\rho,\Phi_n) \leq I_m(\rho,\Phi), \quad \forall n=1,2,\ldots, \text{ and } \liminf_{n\to+\infty} I_m(\rho,\Phi_n) \geq I_m(\rho,\Phi)$$

by monotonicity of relative entropy $H(\cdot\|\cdot)$ and lower semicontinuity of the function $\Phi \mapsto I_m(\rho,\Phi)$, we have

$$I_m(\rho, \Phi) = \sup_n I_m(\rho, \Phi_n).$$

This and lower semicontinuity of the $(\rho, \Phi) \to I_m(\rho, \Phi)$ imply the assertion of the proposition. □

Lemma 11.1.5. *Let $\Phi \in \mathfrak{QC}(A, B)$ be an arbitrary quantum channel from system A to system B and $(\Upsilon_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(B)$ be a sequence of quantum channels from system B to itself that converges strongly to the identity channel $\mathfrak{I}_B : \mathcal{S}(\mathbb{H}_B) \to \mathcal{S}(\mathbb{H}_B)$. Let $(\rho_n)_{n=1}^{+\infty}$ be a sequence of states in $\mathcal{S}(\mathbb{H}_A)$ converging to a state $\rho_0 \in \mathcal{S}(\mathbb{H}_A)$ such that $\lambda_n \rho_n \leq \rho_0$ for some sequence of real numbers $(\lambda_n)_{n=1}^{+\infty}$ converging to 1. Then*

$$\lim_{n \to +\infty} I_m(\rho_n, \Upsilon_n \circ \Phi) = I_m(\rho_0, \Phi).$$

Proof. It follows from the inequality $\lambda_n \rho_n \leq \rho_0$ that $\rho_0 = \lambda_n \rho_n + (1 - \lambda_n)\sigma_n$, where σ_n is a state in $\mathcal{S}(\mathbb{H}_A)$. Hence, concavity and nonnegativity of the mutual information and the first chain rule implies the inequality

$$\lambda_n I_m(\rho_n, \Upsilon_n \circ \Phi) \leq I_m(\rho_0, \Upsilon_n \circ \Phi) \leq I_m(\rho_0, \Phi),$$

showing that

$$\limsup_{n \to +\infty} I_m(\rho_n, \Upsilon_n \circ \Phi) \leq \limsup_{n \to +\infty} \frac{I_m(\rho_0, \Phi)}{\lambda_n} = I_m(\rho_0, \Phi).$$

This and lower semicontinuity of the map $(\rho, \Phi) \to I_m(\rho, \Phi)$ imply

$$\limsup_{n \to +\infty} I_m(\rho_n, \Upsilon_n \circ \Phi) \leq I_m(\rho_0, \Phi) \leq \liminf_{n \to +\infty} I_m(\rho_n, \Upsilon_n \circ \Phi).$$

The assertion of the lemma follows. □

The proof of the next lemma follows easily from the convergence in $\|\cdot\|_1$-norm and properties of partial trace.

Lemma 11.1.6. *Let $\mathbb{H}$ and $\mathbb{K}$ be two finite-dimensional Hilbert spaces such that* $\dim(\mathbb{H}) = \dim(\mathbb{K})$. *For an arbitrary pure state $\omega_0 \in \mathcal{S}(\mathbb{H} \otimes \mathbb{K})$ and an arbitrary sequence $(\rho_k)_{k=1}^{+\infty}$ of states in $\mathcal{S}(\mathbb{H})$ converging to the state $\rho_0 = \mathrm{tr}_{\mathbb{K}}[\omega_0]$, there exists a sequence $(\omega_k)_{k=1}^{+\infty}$ of pure states in $\mathcal{S}(\mathbb{H} \otimes \mathbb{K})$ converging to the state ω_0 such that $\rho_k = \mathrm{tr}_{\mathbb{K}}[\omega_k]$ for all k.*

The following proposition due to Shirokov [150] investigates conditions under which the map $(\rho, \Phi) \mapsto I_m(\rho, \Phi)$ from $\mathcal{A} \times \mathfrak{QC}(A, B)$ to $[-\infty, +\infty]$, where $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$ is continuous under an appropriate product topology.

Proposition 11.1.7. *The map $(\rho, \Phi) \mapsto I_m(\rho, \Phi)$ has the following continuity properties:*

1. *Continuity of the von Neumann entropy $H(\cdot)$ on a set $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$ implies continuity of the map $(\rho, \Phi) \mapsto I_m(\rho, \Phi)$ on the set $\mathcal{A} \times \mathfrak{QC}(A, B)$. Specifically,*

$$\lim_{n\to+\infty} H(\rho_n) = H(\rho_0) < +\infty \Rightarrow \lim_{n\to+\infty} I_m(\rho_n, \Phi_n) = I_m(\rho_0, \Phi_0) < +\infty,$$

for arbitrary sequence $(\Phi_n)_{n=1}^{+\infty}$ of channels strongly converging to a channel Φ_0.

2. *Local continuity of the function $(\rho, \Phi) \mapsto H(\Phi(\rho))$ implies local continuity of the function $(\rho, \Phi) \mapsto I_m(\rho, \Phi)$. Specifically,*

$$\begin{aligned}\lim_{n\to+\infty} H(\Phi_n(\rho_n)) &= H(\Phi_0(\rho_0)) < +\infty \\ &\Rightarrow \lim_{n\to+\infty} I_m(\rho_n, \Phi_n) = I_m(\rho_0, \Phi_0) < +\infty,\end{aligned}$$

for arbitrary sequences $(\rho_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H}_A)$ and $(\Phi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A, B)$ converging respectively to a state ρ_0 and to a channel Φ_0.

3. *Local continuity of the function $(\rho, \Phi) \mapsto I_m(\rho, \Phi)$ implies local continuity of the map $(\rho, \Phi) \mapsto I_m(\rho, \Psi \circ \Phi)$ for any channel $\Psi \in \mathfrak{QC}(B, C)$. That is,*

$$\begin{aligned}\lim_{n\to+\infty} I_m(\rho_n, \Phi_n) &= I_m(\rho_0, \Phi_0) < +\infty \\ &\Rightarrow \lim_{n\to+\infty} I_m(\rho_n, \Psi \circ \Phi_n) = I_m(\rho_0, \Psi \circ \Phi_0) < +\infty\end{aligned}$$

for arbitrary channel $\Psi \in \mathfrak{QC}(B, C)$ and for arbitrary sequences $(\rho_n)_{n=1}^{+\infty} \subset \mathcal{S}(\mathbb{H}_A)$ and $(\Phi_n)_{n=1}^{+\infty} \subset \mathfrak{QC}(A, B)$ converging respectively to a state ρ_0 and to a channel Φ_0.

Proof. Since the strong convergence of a sequence $(\Phi_n)_{n=1}^{+\infty}$ to a channel Φ_0 implies the strong convergence of the sequence $(\Phi_n \otimes \mathfrak{I}_R)_{n=1}^{+\infty}$ (see Lemma 6.6.4) to the channel $\Phi_0 \otimes \mathfrak{I}_R$ assertions 1 and 2 follow immediately. □

11.1.1 Finite-dimensional case

In this subsection, we consider special results of mutual information $I_m(\rho, \Phi)$ for finite-dimensional Hilbert spaces $\mathbb{H}_A$, $\mathbb{H}_B$ and $\mathbb{H}_E$. The presentation of this subsection is based on the results obtained in Cerf and Adami [19] and Holevo and Shirokov [82].

Let $\mathbf{V} : \mathbb{H}_A \to \mathbb{H}_{BE} := \mathbb{H}_B \otimes \mathbb{H}_E$ be an isometric operator and let $\mathbf{V}^* : \mathbb{H}_{BE} \to \mathbb{H}_A$ be its adjoint operator. Then it can be shown via Stinespring representation (see Subsection 4.3) that the a quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and its corresponding complementary channel $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ (see Definition 5.7.1) can be expressed as

$$\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}\rho\mathbf{V}^*] \quad \text{and} \quad \hat{\Phi}(\rho) = \mathrm{tr}_B[\mathbf{V}\rho\mathbf{V}^*], \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A), \tag{11.7}$$

where $\mathrm{tr}_X[\cdot] = \mathrm{tr}_{\mathbb{H}_X}[\cdot]$, define completely positive trace-preserving maps. That is, Φ is a quantum channel from the input system A to output system B and $\hat{\Phi}$ is a quantum channel from the input system A to the environment E. The channels Φ and $\hat{\Phi}$ are called mutually complementary. The construction of Φ and $\hat{\Phi}$ above can be extended to the infinite-dimensional case without changes.

In the following, the identity operator on $\mathbb{H}_X$ and the identity transformation of the set $\mathcal{S}(\mathbb{H}_X)$ will be denoted by $\mathbf{I}_X$ and $\mathfrak{I}_X$, respectively.

Let $\rho = \rho_A$ be an input state in the space $\mathbb{H}_A$, and ρ_B and ρ_E be results of the action of the channels Φ and $\hat{\Phi}$ on the state ρ_A, respectively. That is,

$$\rho_B = \Phi(\rho_A) = \mathrm{tr}_E[\mathbf{V}\rho_A\mathbf{V}^*] \quad \text{and} \quad \rho_E = \tilde{\Phi}(\rho_A) = \mathrm{tr}_B[\mathbf{V}\rho_A\mathbf{V}^*].$$

We have the following results.

Lemma 11.1.8. *Assume that* $\dim(\mathbb{H}_A) < +\infty$ *and* $\dim(\mathbb{H}_B) < +\infty$. *Let* $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ *and* $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ *be mutual quantum channels defined by* (11.7) *and* $\rho = \rho_A$ *be an input state in the space* $\mathbb{H}_A$, $\rho_B = \Phi(\rho_A)$, *and* $\rho_E = \hat{\Phi}(\rho_A)$. *Then*

$$I_m(\rho,\Phi) = H(\rho) + H(\Phi(\rho)) - H(\hat{\Phi}(\rho)), \quad \forall\rho \in \mathcal{S}(\mathbb{H}_A) \tag{11.8}$$

and

$$I_m(\rho,\Phi) = H(\rho) + H(\Phi(\rho)) - H((\Phi \otimes \mathfrak{I}_E)(|\psi_{AE}\rangle\langle\psi_{AE}|)), \quad \forall\rho \in \mathcal{S}(\mathbb{H}_A), \tag{11.9}$$

where E is an environment system and ψ_{AE} is the purification of the state ρ.

Proof. We first note from the definition of mutual information $I_m(\rho,\Phi)$ (see Definition 11.1.1, where we write $R = E$) that

$$\begin{aligned} H((\Phi \otimes \mathfrak{I}_E)(|\varphi_\rho\rangle_{AE}\langle\varphi_\rho|)\|\Phi(\rho) \otimes \rho) &= H(\rho_{BE}\|\rho_B \otimes \rho_E) \\ &= \mathrm{tr}[\rho_{BE}\log(\rho_{BE}) - \log(\rho_B \otimes \rho_E)] \\ &= -H(\rho_{BE}) + H(\rho_B) + H(\rho_E). \end{aligned}$$

This implies that for finite-dimensional spaces $\mathbb{H}_A$, $\mathbb{H}_B$ and $\mathbb{H}_E$ and Φ and $\hat{\Phi}$ defined by (11.7), (11.8) and (11.9) hold. This proves the lemma. □

Based on the same argument as above, we have an analogous characteristic for the complementary channel,

$$I_m(\rho,\hat{\Phi}) = H(\rho_A) + H(\rho_E) - H(\rho_B) = H(\rho_R) + H(\rho_E) - H(\rho_{ER}).$$

We therefore have a fundamental identity

$$I_m(\rho,\Phi) + I_m(\rho,\hat{\Phi}) = 2H(\rho). \tag{11.10}$$

11.1.2 Infinite-dimensional case

We consider complementary channels Φ and $\hat{\Phi}$ in infinite-dimensional spaces $\mathbb{H}_A$, $\mathbb{H}_B$ and $\mathbb{H}_R$ defined by

$$\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}\rho\mathbf{V}^*] \quad \text{and} \quad \hat{\Phi}(\rho) = \mathrm{tr}_B[\mathbf{V}\rho\mathbf{V}^*], \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A), \tag{11.11}$$

where $\mathbf{V} : \mathbb{H}_A \to \mathbb{H}_{BE}$ is an isometric operator and let $\mathbf{V}^* : \mathbb{H}_{BE} \to \mathbb{H}_A$ be its adjoint operator.

The following theorem, due originally to Holevo and Shirokov [82], relates $I_m(\rho, \Phi)$ and $I_m(\phi, \hat{\Phi})$.

Theorem 11.1.9. *Let* $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ *and* $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ *be mutual channels defined by* (11.11). *For an arbitrary state* $\rho \in \mathcal{S}(\mathbb{H}_A)$, *we have the relation*

$$I_m(\rho, \Phi) + I_m(\rho, \hat{\Phi}) = 2H(\rho). \tag{11.12}$$

Proof. Let $\{|h_i\rangle_E\}_{i=1}^{+\infty}$ be an orthonormal basis in the space $\mathbb{H}_E$. Then

$$\mathbf{V}\varphi = \sum_{i=1}^{+\infty} (\mathbf{V}_i\varphi) \otimes |h_i\rangle_E,$$

where $\mathbf{V}_i : \mathbb{H}_A \to \mathbb{H}_B, i = 1, 2, \ldots$ is a sequence bounded operator satisfying the condition $\sum_{i=1}^{+\infty} \mathbf{V}_i^*\mathbf{V}_i = \mathbf{I}_A$. The channel Φ has the Kraus representation $\Phi(\rho) = \sum_{i=1}^{+\infty} \mathbf{V}_i\rho\mathbf{V}_i^*$, and its complementary channel $\hat{\Phi}$ has the representation $\hat{\Phi} = \sum_{i,j=1}^{+\infty} (\mathrm{tr}_B[\mathbf{V}\rho\mathbf{V}^*]))|h_i\rangle_E\langle h_j|$. Let $\rho = \sum_{i=1}^{m} \lambda_i|e_i\rangle_A\langle e_i|$ be a finite-rank quantum state in $\mathcal{S}(\mathbb{H}_A)$, and $\hat{\rho}$ be its purification on $\mathcal{S}(\mathbb{H}_{AR})$. Consider a sequence of linear completely positive trace-nonincreasing maps $(\Phi_n)_{n=1}^{+\infty}$, where $\Phi_n : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is defined as $\Phi_n(\rho) = \sum_{i=1}^{n} \mathbf{V}_i\rho\mathbf{V}_i^*$. The sequence $(\Phi_n)_{n=1}^{+\infty}$ strongly and monotonously converges to Φ, i. e., $\Phi_n(\rho) \le \Phi_{n+1}(\rho)$ for all n and $\rho \in \mathcal{S}(\mathbb{H}_A)$ and $\lim_{n\to+\infty} \Phi_n(\rho) = \Phi(\rho)$ for all $\rho \in \mathcal{S}(\mathbb{H}_A)$.

Since $(\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho})$ is a finite-rank quantum state in $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{H}_R)$, we have

$$\begin{aligned} X_n &= H\big((\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho})\|\Phi(\rho) \otimes \rho\big) \\ &= -\mathrm{tr}\big[\eta\big((\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho})\big)\big] - \mathrm{tr}[(\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho}) \log\big(\Phi(\rho) \otimes \rho\big)] + R_n \end{aligned}$$

where $R_n = 1 - \mathrm{tr}[\Phi_n(\rho)] \to 0$ as $n \to +\infty$. By Lemma 8.2.3, we have $\lim_{n\to+\infty} X_n = I_m(\rho, \Phi)$. Since

$$(\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho}) = \mathrm{tr}_E\big[(\mathbf{I}_B \otimes \mathbf{P}_n \otimes \mathbf{I}_R) \cdot (\mathbf{V} \otimes \mathbf{I}_R)\hat{\rho}(\mathbf{V}^\dagger \otimes \mathbf{I}_R) \cdot (\mathbf{I}_B \otimes \mathbf{P}_n \otimes \mathbf{I}_R)\big],$$

where $\mathbf{P}_n = \sum_{i=1}^{n} |h_i\rangle_E\langle h_i|$ is a finite-dimensional projector on $\mathbb{H}_E$, and the partial trace is taken in the space $\mathbb{H}_B \otimes \mathbb{H}_E \otimes \mathbb{H}_R$. The operator $(\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho})$ is isomorphic to the operator

$$\tilde{\Phi}_n(\rho) = \mathrm{tr}_{BR}[(\mathbf{I}_B \otimes \mathbf{P}_n \otimes \mathbf{I}_R) \cdot (\mathbf{V} \otimes \mathbf{I}_R) \cdot \hat{\rho} \cdot (\mathbf{V}^\dagger \otimes \mathbf{I}_R)(\mathbf{I}_B \otimes \mathbf{P}_n \otimes \mathbf{I}_R)],$$

where $\tilde{\Phi}(\rho) = \mathbf{P}_n \tilde{\Phi}(\cdot)\mathbf{P}_n$ is the quantum operation complementary to Φ_n. Thus, $\mathrm{tr}[\eta(\Phi_n \otimes \mathfrak{I}_R(\hat{\rho}))] = \mathrm{tr}[\eta(\tilde{\Phi}_n(\rho))]$. By using the property of logarithm

$$\log(\rho \otimes \sigma) = \log(\rho) \otimes \mathbf{I} + \mathbf{I} \otimes \log(\sigma)$$

and noting that $\Phi_n(\cdot) \le \Phi(\cdot)$, we obtain

$$\begin{aligned}
&- \mathrm{tr}[(\Phi_n \otimes \mathfrak{I}_R(\hat{\rho}))(\log(\Phi(\rho) \otimes \rho))] \\
&\quad = - \mathrm{tr}[(\Phi_n \otimes \mathfrak{I}_R(\hat{\rho}))(\log(\Phi(\rho)) \otimes \mathbf{I}_R)] - \mathrm{tr}[(\Phi_n \otimes \mathfrak{I}_R(\hat{\rho}))(\mathbf{I}_B \otimes \log(\rho))] \\
&\quad = - \mathrm{tr}[\Phi_n(\rho) \log(\Phi(\rho))] - \mathrm{tr}[(\mathrm{tr}_B(\Phi_n \otimes \mathfrak{I}_R(\hat{\rho})) \log(\rho)].
\end{aligned}$$

Consider the quantity

$$\begin{aligned}
Y_n &= H((\tilde{\Phi}_n \otimes \mathfrak{I}_R)(\hat{\rho}) \| \Phi_n(\rho) \otimes \rho) \\
&= - \mathrm{tr}[\eta((\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho}))] - \mathrm{tr}[(\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho}) \log(\Phi_n(\rho) \otimes \rho)].
\end{aligned}$$

Then $\lim_{n\to\infty} Y_n = I_m(\rho, \tilde{\Phi})$. Similar to the computation of X_n, we obtain

$$\mathrm{tr}[\eta((\tilde{\Phi}_n \otimes \mathfrak{I}_R)(\hat{\rho}))] = \mathrm{tr}[\eta(\Phi_n(\rho))]$$

and

$$\begin{aligned}
&- \mathrm{tr}[((\tilde{\Phi}_n \otimes \mathfrak{I}_R)(\hat{\rho}))(\log(\tilde{\Phi}_n(\rho) \otimes \rho))] \\
&\quad = - \mathrm{tr}[((\tilde{\Phi} \otimes \mathfrak{I}_R)(\hat{\rho}))(\log(\tilde{\Phi}_n(\rho)) \otimes \mathbf{I}_R)] - \mathrm{tr}[(\tilde{\Phi}_n \otimes \mathfrak{I}_R)(\hat{\rho})(\mathbf{I}_E \otimes \log(\rho))] \\
&\quad = \mathrm{tr}[\eta(\tilde{\Phi}_n(\rho))] - \mathrm{tr}[\mathrm{tr}_E[(\tilde{\Phi}_n \otimes \mathfrak{I}_R)(\hat{\rho})] \log(\rho)].
\end{aligned}$$

We next want to show that $\lim_{n\to\infty}(X_n + Y_n) = 2H(\rho)$. By the definition of the relative entropy, we have

$$\begin{aligned}
X_n + Y_n &= - \mathrm{tr}[\Phi_n(\rho) \log(\Phi(\rho))] - \mathrm{tr}[\eta(\Phi_n(\rho))] + C_n + D_n + R_n \\
&= H(\Phi_n(\rho) \| \Phi(\rho)] + C_n + D_n,
\end{aligned}$$

where

$$C_n = - \mathrm{tr}[\mathrm{tr}_B[(\Phi_n \otimes \mathfrak{I}_R)(\hat{\rho})] \log(\rho)] \quad \text{and} \quad D_n = - \mathrm{tr}[\mathrm{tr}_E[(\tilde{\Phi}_n \otimes \mathfrak{I}_R)(\hat{\rho})] \log(\rho)].$$

We next prove $\lim_{n\to+\infty} C_n = \lim_{n\to+\infty} D_n = H(\rho)$. By noting that

$$\mathrm{tr}_B[\Phi_n \otimes \mathfrak{I}_R(\hat{\rho})] = \sum_{i,j=1}^{m} \sqrt{\lambda_i \lambda_j}\, \mathrm{tr}[\Phi(|e_i\rangle\langle e_j|)|e_i\rangle\langle e_j|],$$

we obtain

$$C_n = \sum_{i=1}^{m} (-\lambda_i \log \lambda_i) \operatorname{tr}[\Phi_n(|e_i\rangle\langle e_i|)],$$

and hence, $\lim_{n\to+\infty} C_n = H(\rho)$, since $\lim_{n\to+\infty} \operatorname{tr}[\Phi_n(|e_i\rangle\langle e_i|)] = 1$. In a similar way, one can prove $\lim_{n\to+\infty} D_n = H(\rho)$. Consequently,

$$\lim_{n\to+\infty} H(\Phi_n(\rho)\|\Phi(\rho)) = 0.$$

Thus, we have $\lim_{n\to+\infty}(X_n + Y_n) = 2H(\rho)$. Since $\lim_{n\to+\infty} X_n = I_m(\rho, \Phi)$ and $\lim_{n\to+\infty} Y_n = I_m(\rho, \tilde{\Phi})$, the assertion of the theorem is proved for finite-rank states. Since the left- and right-hand sides of relation (11.12) are concave lower semicontinuous nonnegative functions, validity of this relation for all states follows because concave lower semicontinuous lower bounded function on a set of quantum states is uniquely determined by its restriction to the set of finite-rank states. This proves the theorem. □

11.2 Coherent information

The coherent information was first introduced by Schumacher and Nielsen [162] and Lloyd [109]. It is an entropic measure of a quantum state ρ and a quantum channel Φ in an attempt to describe how much of the quantum information in the state will remain after the state goes through the channel. In this sense, it is intuitively similar to the mutual information of classical information theory.

We first give a definition of coherent information for the finite-dimensional case below (see, e. g., Wilde [178] and Nielson and Chung [116] for the definition and general properties).

Definition 11.2.1 (Finite dimensions). Assume that $\dim(\mathbb{H}_A) < \infty$, $\dim(\mathbb{H}_B) < \infty$ and $\dim(\mathbb{H}_E) < \infty$, the coherent information $I_c(\rho, \Phi)$ of quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ at (input) quantum state $\rho \in \mathcal{S}(\mathbb{H}_A)$ is defined by

$$I_c(\rho, \Phi) = H(\Phi(\rho)) - H(\rho_E) = H(\Phi(\rho)) - H(\rho_{AR}), \tag{11.13}$$

where reference system $\mathbb{H}_R \cong \mathbb{H}_A$ (a copy of $\mathbb{H}_A$) is a reference system and $\psi_{AR} \in \mathcal{S}(\mathbb{H}_{AR})$ is the purification vector for the state ρ_A, and $\rho_{AR} = (\Phi \otimes \mathfrak{I}_R)(|\psi_{AR}\rangle\langle\psi_{AR}|)$.

Since in the infinite-dimensional case the right-hand side in Definition 11.2.1 of the coherent information $I_c(\rho, \Phi)$ can be indefinite (i. e., in the form of $\infty - \infty$) even for a state ρ with finite entropy $H(\rho)$ and finite mutual information $I_m(\rho, \Phi)$ for such state ρ and any channel Φ. Following Holevo and Shirokov [82], we define the coherent information for an infinite-dimensional quantum channel as follows to avoid such a pitfall.

Definition 11.2.2 (Infinite dimensions). Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel, and ρ be a state in $\mathcal{S}(\mathbb{H}_A)$ with finite entropy. The coherent information of the channel Φ at the state ρ is defined as follows:

$$I_c(\rho,\Phi) = I_m(\rho,\Phi) - H(\rho). \tag{11.14}$$

Remark 11.1. Following from Theorem 11.1.9 that $I_m(\rho,\Phi) + I_m(\rho,\hat{\Phi}) = 2H(\rho)$ and the above definition of $I_c(\rho,\Phi)$, we immediately have the following relation:

$$I_c(\rho,\Phi) + I_c(\rho,\hat{\Phi}) = 0, \tag{11.15}$$

where $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ is the complementary channel of Φ.

The coherent information $I_c(\rho,\Phi)$ of the channel Φ at the state ρ has the following basic properties, which follow immediately by its definition and from the basic properties of mutual information $I_m(\rho,\Phi)$ (see Proposition 11.1.4) and the relation between $I_m(\rho,\Phi)$ and $I_c(\rho,\Phi)$ as defined in (11.14).

The following three results can be found in Holevo and Shirokov [82].

Proposition 11.2.3. *The map* $(\rho,\Phi) \mapsto I_c(\rho,\Phi)$ *is nonnegative and lower semicontinuous on the set* $\mathcal{S}(\mathbb{H}_A) \times \mathfrak{QC}(A,B)$. *It has the following properties:*
1. Concavity *in* ρ*:* $I_c(p\rho_1 + (1-p)\rho_2,\Phi) \geq pI_c(\rho_1,\Phi) + (1-p)I_c(\rho_2,\Phi)$ *for all* $0 \leq p \leq 1$, $\rho_1,\rho_2 \in \mathcal{S}(\mathbb{H}_A)$ *and* $\Phi \in \mathfrak{QC}(A,B)$*;*
2. Convexity *in* Φ*:* $I_c(\rho,p\Phi_1 + (1-p)\Phi_2) \leq pI_c(\rho,\Phi_1) + (1-p)I_c(\rho,\Phi_2)$ *for all* $0 \leq p \leq 1$, $\rho \in \mathcal{S}(\mathbb{H}_A)$ *and* $\Phi_1,\Phi_2 \in \mathfrak{QC}(A,B)$*;*
3. The first chain rule*: for arbitrary channels* $\Phi \in \mathfrak{QC}(A,B)$ *and* $\Psi \in \mathfrak{QC}(B,C)$ *the inequality* $I_c(\rho,\Psi\circ\Phi) \leq I_c(\rho,\Phi)$ *holds for any* $\rho \in \mathcal{S}(\mathbb{H}_A)$*;*
4. The second chain rule*: for arbitrary channels* $\Phi \in \mathfrak{QC}(A,B)$ *and* $\Psi \in \mathfrak{QC}(B,C)$*, the inequality* $I_c(\rho,\Psi\circ\Phi) \leq I_c(\Phi(\rho),\Psi)$ *holds for any* $\rho \in \mathcal{S}(\mathbb{H}_A)$*;*
5. Subadditivity*: for arbitrary channels* $\Phi \in \mathfrak{QC}(A,B)$ *and* $\Psi \in \mathfrak{QC}(C,D)$ *the inequality*

$$I_c(\omega,\Phi\otimes\Psi) \leq I_c(\omega_A,\Phi) + I_c(\omega_C,\Psi)$$

holds for any $\omega \in \mathcal{S}(\mathbb{H}_A \otimes \mathbb{H}_C)$*, where* $\omega_A = \mathrm{tr}_C[\omega]$ *and* $\omega_C = \mathrm{tr}_A[\omega]$.

Proposition 11.2.4. *For an arbitrary quantum channel* $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$*, the functions* $\rho \to I_m(\rho,\Phi)$ *and* $\rho \to I_c(\rho,\Phi)$ *are continuous on any subset* $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$ *on which the von Neumann entropy* $H(\cdot)$ *is continuous.*

Proof. By Proposition 11.1.4, the functions $\rho \to I_m(\rho,\Phi)$ and $\rho \to I_c(\rho,\Phi)$ are lower semicontinuous, while by Theorem 11.1.9, $I_m(\rho,\Phi) + I_c(\rho,\Phi) = H(\rho)$, which is continuous on the set $\mathcal{A}$ by the condition. Hence, these functions are continuous on the set $\mathcal{A}$.

The function $\rho \to I_c(\rho, \Phi)$ is continuous on the set $\mathcal{A}$ as a difference of two functions that are continuous on this set. This proves the proposition. □

Proposition 11.2.5. *Let $(\Phi_n)_{n=1}^{+\infty}$ be a sequence of channels in $\mathfrak{QC}(A, B)$ and strongly converging to a channel Φ_0. Assume that there exists a sequence $(\hat{\Phi}_n)_{n=1}^{+\infty}$ of channels in $\mathfrak{QC}(A, E)$ strongly converging to a channel $\hat{\Phi}_0$ such that $(\Phi_n, \hat{\Phi}_n)$ is a complementary pair for each $n = 0, 1, 2, \ldots$. Then the relations*

$$\lim_{n\to+\infty} H(\rho_n) = \lim_{n\to+\infty} I_m(\rho_n, \Phi_n) = I_m(\rho_0, \Phi_0) \tag{11.16}$$

and

$$\lim_{n\to+\infty} I_c(\rho_n, \Phi_n) = I_c(\rho_0, \Phi_0) \tag{11.17}$$

hold for any sequence $(\rho_n)_{n=1}^{+\infty}$ of states in $\mathcal{S}(\mathbb{H}_A)$ converging to the state ρ_0 such that $\lim_{n\to+\infty} H(\rho_n) = H(\rho_0) < +\infty$.

Proof. This follows from the continuity of the $I_m(\cdot,\cdot)$ and $I_c(\cdot,\cdot)$ from $\mathcal{S}(\mathbb{H}_A) \times \mathfrak{QC}(A, B)$. □

In the following, we use the following Holevo channel output χ-function $\chi_\Phi(\mu)$ defined by

$$\chi_\Phi(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho)\|\Phi(\overline{\rho}(\mu)))\mu(d\rho),$$

where $\overline{\rho}(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu(d\rho)$ is the barycenter of the measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ (see Definition 12.1.2).

Lemma 11.2.6. *Let $(\Phi_n)_{n=1}^{+\infty}$ be a sequence of quantum operations, strongly converging to the channel Φ_0. The relations*

$$\lim_{n\to+\infty} \overline{\mathrm{co}}(H_{\Phi_n}(\rho)) = \overline{\mathrm{co}}(H_{\Phi_0}(\rho)) \text{ and } \lim_{n\to+\infty} \chi_{\Phi_n}(\rho) = \chi_{\Phi_0}(\rho),$$

hold for any state ρ in $\mathcal{S}(\mathbb{H}_A)$ in the following cases:
(A) $\Phi_n(\cdot) = \mathbf{P}_n\Phi_0(\cdot)\mathbf{P}_n$ *for some sequence $(\mathbf{P}_n)_{n=1}^{+\infty}$ of projectors in $\mathfrak{B}(\mathbb{H}_B)$,*
(B) $\Phi_n(\rho) \le \Phi_0(\rho)$ *for all ρ in $\mathcal{S}(\mathbb{H}_A)$ (in the operator order).*

Proof. (A) For arbitrary state ρ in $\mathcal{S}(\mathbb{H}_A)$, the monotonicity of the relative entropy (see (9.2)) implies $\overline{\mathrm{co}}(H_{\Phi_n}(\rho)) \le \overline{\mathrm{co}}(H_{\Phi_0}(\rho))$ and $\chi_{\Phi_n}(\rho) \le \chi_{\Phi_0}(\rho)$, correspondingly. Hence, the limit relations in the lemma follow from the fact that the functions $(\rho, \Phi) \mapsto \chi_\Phi(\rho)$ and $(\rho, \Phi) \mapsto \overline{\mathrm{co}}H_\Phi(\rho)$ on the set $\mathcal{S}(\mathbb{H}_A) \times \mathfrak{QC}(A, B)$.

(B) For arbitrary state ρ in $\mathcal{S}(\mathbb{H}_A)$, the following inequality:

$$H(\mathbf{A}) + H(\mathbf{B} - \mathbf{A}) \leq H(\mathbf{B}) \leq H(\mathbf{A}) + H(\mathbf{B} - \mathbf{A}) + \mathrm{tr}\left[\mathbf{B}h_2\left(\frac{\mathrm{tr}[\mathbf{B}]}{\mathrm{tr}[\mathbf{A}]}\right)\right]$$

(where $\mathbf{A}, \mathbf{B} \in \mathfrak{T}_1(\mathbb{H}_A)$, $\mathbf{A} \leq \mathbf{B}$ and $h_2 = \eta(x) + \eta(1 - x)$) implies that

$$\overline{\mathrm{co}}(H_{\Phi_n}(\rho)) \leq \overline{\mathrm{co}}(H_{\Phi_0}(\rho)) \quad \text{and} \quad H_{\Phi_n}(\rho) \leq H_{\Phi_0}(\rho) + (\mathrm{tr}[\Phi_n(\rho))] + h_2(\mathrm{tr}[\Phi_n(\rho)]),$$

correspondingly. Hence, the limit relations in the proposition follows. This proves the lemma. □

The following result provides an alternative expression for the coherent information of the channel Φ at the state ρ, $I_c(\rho, \Phi)$, with $H(\rho) < +\infty$ and $H(\Phi(\rho)) < +\infty$. The proof can be obtained by using the relation of this quantity with the secret classical capacity of a channel mentioned in Schumacher and Westmoreland [139] and, therefore, is omitted here.

Proposition 11.2.7. *Let $\rho = \sum_i p_i\rho_i \in \mathcal{S}(\mathbb{H}_A)$ and $\Phi \in \mathfrak{QC}(A, B)$ be such that $H(\rho) < +\infty$ and $H(\Phi(\rho)) < +\infty$. Let $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ be the complementary channel of Φ. Then*

$$I_c(\rho, \Phi) = \chi_\Phi(\rho) - \chi_{\hat{\Phi}}(\rho), \tag{11.18}$$

where

$$\chi_\Phi(\rho) = \sup \sum_i p_i H\left(\Phi(\rho_i) \| \Phi\left(\sum_i p_i\rho_i\right)\right)$$

and

$$\chi_{\hat{\Phi}}(\rho) = \sup \sum_i p_i H\left(\hat{\Phi}(\rho_i) \| \hat{\Phi}\left(\sum_i p_i\rho_i\right)\right),$$

in which the supremum above is taken over all convex decompositions $\rho = \sum_i p_i\rho_i$, $\rho_i \in \mathcal{S}(\mathbb{H}_A)$.

Let $\mathfrak{B}_{\mathrm{seq}}(\mathbb{H}_A, \mathbb{H}_B)$ be the set of all sequences $\vec{\mathbf{V}} = (\mathbf{V}_i)_{i=1}^{+\infty}$ of bounded linear operators from $\mathbb{H}_A$ to $\mathbb{H}_B$ such that $\sum_{i=1}^{+\infty} \mathbf{V}_i^*\mathbf{V}_i = \mathbf{I}_A$, endowed with the topology of coordinatewise strong operator convergence. That is, $\vec{\mathbf{V}}^{(n)} = (\mathbf{V}_i^{(n)})_{i=1}^{+\infty} \in \mathfrak{B}_{\mathrm{seq}}(\mathbb{H}_A, \mathbb{H}_B)$ for $n = 1, 2, \ldots$, is said to converge to $\vec{\mathbf{V}} = (\mathbf{V}_i)_{i=1}^{+\infty} \in \mathfrak{B}_{\mathrm{seq}}(\mathbb{H}_A, \mathbb{H}_B)$ as $n \to +\infty$ if

$$\lim_{n\to+\infty} \|\mathbf{V}_i^{(n)} - \mathbf{V}_i\|_\infty = 0, \quad \forall i = 1, 2, \ldots.$$

We have the following corollary.

Corollary 11.2.8. *For an arbitrary subset $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H}_A)$ on which the von Neumann entropy $H(\cdot)$ is continuous, the functions*

$$(\rho, \vec{\mathbf{V}}) \mapsto I_m(\rho, \Phi[\vec{\mathbf{V}}]), \quad (\rho, \vec{\mathbf{V}}) \mapsto I_c(\rho, \Phi[\vec{\mathbf{V}}]),$$

and

$$(\rho, \vec{\mathbf{V}}) \mapsto \sum_{i=1}^{+\infty} H(\mathbf{V}_i \rho \mathbf{V}^*),$$

where $\Phi[\vec{\mathbf{V}}](\cdot) = \sum_{i=1}^{+\infty} \mathbf{V}_i(\cdot)\mathbf{V}_i^*$, are continuous on the set $\mathcal{A} \times \mathfrak{B}_{\mathrm{seq}}(\mathbb{H}_A, \mathbb{H}_B)$.

Proof. By Proposition 11.2.5, continuity of the first two functions follows from continuity of the maps

$$\vec{\mathbf{V}} \mapsto \Phi[\vec{\mathbf{V}}] \in \mathfrak{QC}(A, B) \quad \text{and} \quad \vec{\mathbf{V}} \mapsto \hat{\Phi}[\vec{\mathbf{V}}] \in \mathfrak{QC}(A, E),$$

where $\hat{\Phi}[\vec{\mathbf{V}}](\cdot) = \sum_{i,j=1}^{+\infty} \mathrm{tr}_B[\mathbf{V}_i(\cdot)\mathbf{V}_j^*](|h_i\rangle_E\langle h_j|)$, and $(|h_i\rangle_E)_{i=1}^{+\infty}$ is an orthonormal basis in $\mathbb{H}_E$. To prove continuity of these maps, it suffices to show that

$$\lim_{n\to+\infty} \Phi[\vec{\mathbf{V}}_n](|\varphi\rangle_A\langle\varphi|) = \Phi[\vec{\mathbf{V}}_0](|\varphi\rangle_A\langle\varphi|) \tag{11.19}$$

and

$$\lim_{n\to+\infty} \hat{\Phi}[\vec{\mathbf{V}}_n](|\varphi\rangle_A\langle\varphi|) = \hat{\Phi}[\vec{\mathbf{V}}_0](|\varphi\rangle_A\langle\varphi|) \tag{11.20}$$

for any sequence $(\vec{\mathbf{V}}^{(n)})_{n=1}^{+\infty} \subset \mathfrak{B}_{\mathrm{seq}}(\mathbb{H}_A, \mathbb{H}_B)$ converging to a vector $\vec{\mathbf{V}}_0 \in \mathfrak{B}_{\mathrm{seq}}(A, B)$ and for any unit vector $\varphi \in \mathbb{H}_A$. Let $\vec{\mathbf{V}}^{(n)} = (\mathbf{V}_i^{(n)})_{i=1}^{+\infty}$ for each $n \geq 0$. Relation 11.19 can be proved by noting that the condition $\sum_{i=1}^{+\infty} \|\mathbf{V}_i^{(n)}\|_\infty^2 = 1$ for all $n \geq 0$ implies

$$\lim_{m\to+\infty} \sup_{n\geq 0} \mathrm{tr}\left[\sum_{i>m} \mathbf{V}_i^{(n)}(|\varphi\rangle_A\langle\varphi|)\mathbf{V}_i^{(n)*}\right] = \lim_{m\to+\infty} \sup_{n\geq 0} \sum_{i>m} \|\mathbf{V}_i^{(n)}|\varphi\rangle_A\|_{\mathbb{H}_B}^2 = 0.$$

Relation 11.20 is easily proved by using the fact that convergence of quantum states in trace norm $\|\cdot\|_1$ is equivalent to weak operator convergence (see Section 2.1).

To prove continuity of the third function, we let $\mathbb{H}_C = \oplus_{i=1}^{+\infty}\mathbb{H}_B^i$, where $\mathbb{H}_B^i = \mathbb{H}_B$ (an identical copy of $\mathbb{H}_B$), and let $\mathbf{U}_i$ be an isometric embedding of $\mathbb{H}_B$ in $\mathbb{H}_C$ such that $\mathbf{U}_i\mathbb{H}_B = \mathbb{H}_B^i$ for each i. For an arbitrary sequence $(\mathbf{V}_i)_{i=1}^{+\infty}$ in $\mathfrak{B}_1(A, B)$, one can take the sequence $(\hat{\mathbf{V}}_i)_{i=1}^{+\infty}$ in $\mathfrak{B}_1(A, C)$, where $\hat{\mathbf{V}}_i = \mathbf{U}_i\mathbf{V}_i$ such that $\mathrm{range}(\hat{\mathbf{V}}_i) \perp \mathrm{range}(\hat{\mathbf{V}}_j)$ for all $i \neq j$. Since the above correspondence is continuous (as a map from $\mathfrak{B}_1(A, B)$ to $\mathfrak{B}_1(A, C)$), the above observation shows continuity of the function

$$(\rho, \vec{\mathbf{V}}) \mapsto I_c(\rho, \hat{\Phi}[\vec{\mathbf{V}}]) = \sum_{i=1}^{+\infty} H(\hat{\mathbf{V}}_i \rho \hat{\mathbf{V}}_i^*) = \sum_{i=1}^{+\infty} H(\mathbf{V}_i \rho \mathbf{V}_i^*)$$

on the set $\mathcal{A} \times \mathfrak{B}_1(A, B)$, where $\hat{\Phi}[\vec{\mathbf{V}}](\cdot) = \sum_{i=1}^{+\infty} \hat{\mathbf{V}}_i(\cdot)\hat{V}_i^*$ and where the first equality follows from the last assertion of Proposition 11.2.5. This proves the corollary. □

As was mentioned in Chapter 9, the quantity $\sum_{i=1}^{+\infty} H(\mathbf{V}_i \rho \mathbf{V}_i^*)$ can be considered as the mean entropy of an a posteriori state in the quantum measurement described by the collection of operators $(\mathbf{V}_i)_{i=1}^{+\infty}$. Corollary 11.2.8 shows that continuity of the entropy $H(\rho)$ of an a priori state ρ implies continuity of the mean entropy of the a posteriori state as a function of a pair (a priori state, measurement) provided that the strong operator topology is used in the definition of convergence of a sequence of measurements. This assertion strengthens an analogous assertion in Example 3 in Shirokov [144], where a stronger topology (the so-called *-strong operator topology) was used in the definition of convergence of a sequence of measurements. Hence, with the use of Corollary 11.2.8, one can strengthen all the assertions in Example 3 in [144] by inserting the strong operator topology in the definition of convergence of a sequence of measurements, which seems more natural in this context.

Consider the channel $\Phi(\cdot) = \mathrm{tr}_E[\cdot] : \mathcal{S}(\mathbb{H}_{BE}) \to \mathcal{S}(\mathbb{H}_B)$ and note (see Kuznetsova [102]) that its complementary channel is $\hat{\Phi}(\cdot) = \mathrm{tr}_B[\cdot] : \mathcal{S}(\mathbb{H}_{BE}) \to \mathcal{S}(\mathbb{H}_E)$. If we write Definition 11.2.2 of coherent information of the channels Φ and $\hat{\Phi}$ at the state ρ_{BE}, we have

$$\begin{aligned} I_c(\rho_{BE}, \mathrm{tr}_E[\cdot]) &= H(\rho_{AB} \| \rho_A \otimes \rho_B) - H(\rho_{BE}) \\ &= H(\rho_{AB} \| \rho_A \otimes \rho_B) - H(\rho_A) = -H(A|B), \\ I_c(\rho_{BE}, \mathrm{tr}_B[\cdot]) &= H(\rho_{AE} \| \rho_A \otimes \rho_E) - H(\rho_{BE}) \\ &= H(\rho_{AE} \| \rho_A \otimes \rho_E) - H(\rho_A) = -H(A|E). \end{aligned}$$

We use the fact that $H(\rho_{BE}) = H(\rho_A)$, since ρ_{ABE} is a pure state. Since coherent information $I_c(\rho, \Phi)$ of any channel Φ at the state ρ takes values in $[-H(\rho), H(\rho)]$, a similar relation for conditional entropy is

$$-H(A) \le H(A|B) \le H(A).$$

Due to the previous result for coherent information, we obtain the important identity

$$H(A|B) + H(A|E) = 0,$$

which occurs in the case when ρ_{ABE} is a pure state of system $\mathbb{H}_A \otimes \mathbb{H}_B \otimes \mathbb{H}_E$.

Recall from Definition 5.6.4 that a channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is perfectly reversible on a state $\rho \in \mathcal{S}(\mathbb{H}_A)$ if there exists a channel $\Psi : \mathcal{S}(\mathbb{H}_B) \to \mathcal{S}(\mathbb{H}_A)$ such that $\Psi \circ \Phi(\sigma) = \sigma$ for all states σ with $\mathrm{supp}(\sigma) \subseteq \mathrm{supp}(\rho)$.

The following result, due originally to Shirokov [148], characterizes a perfect reversibility of a channel Φ in terms of its mutual information $I_m(\rho, \Phi)$ and coherent information $I_c(\rho, \Phi)$.

Proposition 11.2.9. *Let $\rho \in \mathcal{S}(\mathbb{H}_A)$ be such that $H(\rho) < +\infty$. A channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is perfectly reversible on the state ρ if and only if one of following equivalent conditions holds:* (i) $I_c(\rho, \Phi) = H(\rho)$ *and* (ii) $I_m(\rho, \tilde{\Phi}) = 0$, *where $\tilde{\Phi}$ is the complimentary channel of Φ.*

Proof. By Theorem 11.1.9 and Definition 11.2.2, we have

$$H(\rho) - I_c(\rho, \Phi) = I_m(\rho, \Phi) \geq 0,$$

where the equality holds if and only if $\rho_{RE} = \rho_R \otimes \rho_E$, since $I_m(\rho, \Phi) = H(\rho_{RE} \| \rho_R \otimes \rho_E)$.

($\Rightarrow$) Assume that the channel Φ is perfectly reversible. Let $\mathbf{V} : \mathbb{H}_A \to \mathbb{H}_{BE}$ be the isometry from the following representation of the channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and its complementary channel $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$;

$$\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}\rho\mathbf{V}^*] \text{ and } \tilde{\Phi}(\rho) = \mathrm{tr}_B[\mathbf{V}\rho\mathbf{V}^*], \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A).$$

Consider the pure state $\rho_{BRE} = |\varphi_{BRE}\rangle\langle\varphi_{BRE}|$, where $|\varphi_{BRE}\rangle = (\mathbf{V} \otimes \mathbf{I}_R)|\varphi_{AR}\rangle$. Since the channel Φ is perfectly reversible, there exists a channel Ψ such that (5.27) holds, and hence,

$$(\Psi \otimes \mathfrak{I}_{RE})(\rho_{BRE}) = \rho_{ARE}.$$

Since ρ_{AR} is a pure state, we have $\rho_{ARE} = \rho_{AR} \otimes \rho_E$. By taking partial traces over the space $\mathbb{H}_A$, we obtain $\rho_{RE} = \rho_R \otimes \rho_E$. Therefore,

$$H(\rho) - I_c(\rho, \Phi) = I_m(\rho, \Phi) = 0.$$

($\Leftarrow$) Consider the vector $|\varphi_{BRE}\rangle = (\mathbf{V} \otimes \mathbf{I}_R)|\varphi_{AR}\rangle$. Then $|\varphi_{BRE}\rangle$ is a purification vector for the state ρ_{RE}. Since $\rho_{RE} = \rho_R \otimes \rho_E$, $|\varphi_{AR}\rangle \otimes |\varphi_{EE}\rangle$ is a purification vector for the state ρ_{RE}, where E is a reference system for the system E. Without loss of generality, we can assume that the Hilbert spaces of the both purifications are infinite-dimensional, so that there exists an isometry $\mathbf{W} : \mathbb{H}_B \to \mathbb{H}_A \otimes \mathbb{H}_E$ such that

$$(\mathbf{I}_{RE} \otimes \mathbf{W})|\varphi_{BRE}\rangle = |\varphi_{AR}\rangle \otimes |\varphi_{EE}\rangle,$$

and, respectively,

$$(\mathbf{I}_{RE} \otimes \mathbf{W})(|\varphi_{BRE}\rangle\langle\varphi_{BRE}|)((\mathbf{I}_{RE} \otimes \mathbf{W}^*) = (|\varphi_{AR}\rangle\langle\varphi_{AR}|) \otimes |\varphi_{EE'}\rangle\langle\varphi_{EE'}|).$$

By taking partial traces over the spaces $\mathbb{H}_E$ and $\mathbb{H}_E$, we obtain the perfect reversibility condition (5.27), where

$$\Psi(\sigma) = \mathrm{tr}_E[\mathbf{W}\sigma\mathbf{W}^*], \quad \forall \sigma \in \mathcal{S}(\mathbb{H}_B).$$

This proves the proposition. □

12 Holevo χ-capacity

In this chapter, we will be concerned with the constrained and unconstrained capacity for sending classical information over one single use of the noisy quantum channel Φ. This capacity will be called an Holevo χ-capacity. The unconstrained and constrained Holevo χ-capacities will be denoted by $C_\chi(\Phi)$ and $C_\chi(\Phi; \mathcal{A})$, respectively. Channel capacities of various types obtained in the next four chapters are expressed in terms of Holevo χ-capacities.

The concept of χ-quantity and χ-capacity in finite dimensions was first created by Holevo [67] (see also Holevo [77]). Recently, these quantities have been extended to the infinite-dimensional setting by Holevo and his collaborators (see, e. g., Holevo [72, 75, 76, 79, 83, 84] and [85]) and, therefore, applicable to infinite-dimensional quantum information theory. The presentation of topics in this chapter is largely based on the results from Holevo and Shirokov [79, 83], Holevo–Shirokov–Werner [84, 85] and Shirokov [142].

12.1 The χ-functions

12.1.1 Input χ-function

In the following, let $\mathbb{H}_A$ and $\mathbb{H}_B$ be the separable complex Hilbert spaces that represent the input quantum system A and the output quantum system B, respectively. Unless otherwise stated, $\mathbb{H}_A$ and $\mathbb{H}_B$ are assumed to be infinite-dimensional.

Consider an arbitrary probability measure μ on the Borel measurable space $(\mathcal{S}(\mathbb{H}_A), \mathcal{B}(\mathcal{S}(\mathbb{H}_A)))$ on $\mathcal{S}(\mathbb{H}_A)$. Again, let $H(\cdot) : \mathcal{S}(\mathbb{H}_A) \to [0, +\infty]$ be the von Neumann channel input entropy and $H(\cdot\|\cdot) : \mathcal{S}(\mathbb{H}_A) \times \mathcal{S}(\mathbb{H}_A) \to [-\infty, +\infty]$ be the relative entropy (see Definition 7.1.1 and Definition 8.1.3, resp.).

We first give a definition of unconstrained input χ-function (or Holevo quantity) $\chi(\cdot) : \mathcal{P}(\mathcal{S}(\mathbb{H}_A)) \to [-\infty, +\infty]$ as follows.

Definition 12.1.1 (Input χ-function). The unconstrained input χ-function of a probability measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$, denoted by $\chi(\mu)$, is defined by

$$\chi(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} H\Big(\rho \Big\| \int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu(d\rho)\Big)\mu(d\rho). \tag{12.1}$$

If $H(\bar{\rho}(\mu)) = H(\int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu(d\rho)) < +\infty$, then it follows from the fact that $H(\rho\|\sigma) = H(\sigma) - H(\rho)$ for $\rho, \sigma \in \mathcal{S}(\mathbb{H}_A)$ that the input χ-function $\chi(\mu)$ can be written as

https://doi.org/10.1515/9783110788105-012

$$\chi(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\rho\|\bar{\rho}(\mu))\mu(d\rho) = H\Big(\int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu(d\rho)\Big) - \int_{\mathcal{S}(\mathbb{H}_A)} H(\rho)\mu(d\rho)$$
$$= H(\bar{\rho}(\mu)) - \int_{\mathcal{S}(\mathbb{H}_A)} H(\rho)\mu(d\rho). \tag{12.2}$$

The *unconstrained input χ-function* is also referred to as an *unconstrained input Holevo quantity* and we often use these different terminologies interchangeably in the following.

In the case where $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ is a discrete ensemble, i. e., $\mu = \{p_i, \rho_i\}$, then the unconstrained input χ-function $\chi(\{p_i, \rho_i\})$ takes the following form:

$$\chi(\{p_i, \rho_i\}) = \sum_i p_i H(\rho_i\|\bar{\rho}(\mu)) = H(\bar{\rho}(\mu)) - \sum_i p_i H(\rho_i)$$
$$= H\Big(\sum_i p_i\rho_i\Big) - \sum_i p_i H(\rho_i). \tag{12.3}$$

Note that the unconstrained input χ-function provides an upper bound for the accessible classical information, which can be obtained by applying a quantum measurement (see Subsection 2.6.1 for the concept of quantum measurement), which will be explored in detail in the next chapter.

12.1.2 Output χ-function

Consider an arbitrary probability measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ as an input generalized ensemble for a quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$, where $\mathbb{H}_A$ and $\mathbb{H}_B$ are the complex Hilbert spaces representing the input system A and the output system B, respectively.

We define the unconstrained output χ-function (or output Holevo quantity), $\chi_\Phi(\mu)$, for the channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ as follows.

Definition 12.1.2 (Unconstrained channel output χ-function).

$$\chi_\Phi(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho)\|\Phi(\bar{\rho}(\mu)))\mu(d\rho), \tag{12.4}$$

where $\bar{\rho}(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu(d\rho)$ is the barycenter or average state of the measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$.

Lemma 12.1.3. *Let $\{p_i, \rho_i\}_{i=1}^m$ be an arbitrary ensemble of m states with the average state $\bar{\rho} = \sum_{i=1}^m p_i\rho_i$ and let $(\bar{\rho}_n)_{n=1}^{+\infty}$ be an arbitrary sequence of states converging to the state $\bar{\rho}$. Then there exists the sequence $(\{p_i^n, \rho_i^n\}_{i=1}^m)_{n=1}^{+\infty}$ of ensembles of m states such that*

$$\lim_{n\to+\infty} p_i^n = p_i, \quad \lim_{n\to+\infty} \rho_i^n = \rho_i, \quad \forall i = 1, 2, \dots, m,$$

and

$$\bar{\rho}_n = \sum_{i=1}^{m} p_i^n \rho_i^n, \quad \forall n = 1, 2, \ldots.$$

Proof. Without loss of generality, we may assume that $p_i > 0$ for $i = 1, 2, \ldots, m$. Let $\mathbb{D} \subseteq \mathbb{H}$ be the support of $\bar{\rho} = \sum_{i=1}^{m} p_i \rho_i$ and $\mathbf{P}$ be the projector onto $\mathbb{D}$. Since $\rho_i \leq \frac{\bar{\rho}}{p_i}$, we have

$$0 \leq \mathbf{A}_i := (\sqrt{\bar{\rho}})^{-1} \rho_i (\sqrt{\bar{\rho}})^{-1} \leq \frac{\mathbf{I}}{p}, \quad i = 1, 2, \ldots, m,$$

where we denote by $(\sqrt{\bar{\rho}})^{-1}$ the generalized Moore–Penrose inverse (see Moore [114] and Penrose [124]) of the operator $\sqrt{\bar{\rho}}$ (equal 0 on the orthogonal complement to $\mathbb{D}$). Consider the sequence $(\mathbf{B}_i^n)_{n=1}^{+\infty}$, where

$$\mathbf{B}_i^n = \sqrt{\bar{\rho}_n} \mathbf{A}_i \sqrt{\bar{\rho}_n} + \sqrt{\bar{\rho}_n} (\mathbf{I} - \mathbf{P}) \sqrt{\bar{\rho}_n}, \quad n = 1, 2, \ldots$$

of operators in $\mathfrak{B}(\mathbb{H})$. Since $\lim_{n \to +\infty} \bar{\rho}_n = \bar{\rho} = \mathbf{P}\bar{\rho}$ in the trace-norm $\| \cdot \|_1$, we have

$$\lim_{n \to +\infty} \mathbf{B}_i^n = \sqrt{\bar{\rho}} \mathbf{A}_i \sqrt{\bar{\rho}} = \rho_i$$

in the weak operator topology. The last equality implies $\mathbf{A}_i \neq \mathbf{O}$. Note that

$$\mathrm{tr}[\mathbf{B}_i^n] = \mathrm{tr}[\mathbf{A}_i \bar{\rho}_n] + \mathrm{tr}[(\mathbf{I}_{\mathbb{H}} - \mathbf{P}) \bar{\rho}_n] < +\infty,$$

and hence,

$$\lim_{n \to +\infty} \mathrm{tr}[\mathbf{B}_i^n] = \mathrm{tr}[\mathbf{A}_i \bar{\rho}] + \mathrm{tr}[(\mathbf{I}_{\mathbb{H}} - \mathbf{P}) \mathbf{P}] = \mathrm{tr}[\rho_i] = 1.$$

Denote by $\rho_i^n = (\mathrm{tr}[\mathbf{B}_i^n])^{-1} \mathbf{B}_i^n$ (a quantum state) and by $p_i^n = p_i \,\mathrm{tr}[\mathbf{B}_i^n]$ (a positive number) for each i, then $\lim_{n \to +\infty} p_i^n = p_i$ and $\lim_{n \to +\infty} \rho_i^n = \rho_i$ in the weak operator topology, and hence, by the result in the trace norm. Moreover,

$$\begin{aligned} \sum_{i=1}^{m} p_i^n \rho_i^n &= \sum_{i=1}^{m} p_i \mathbf{B}_i^n \\ &= \sqrt{\bar{\rho}_n} (\sqrt{\rho})^{-1} \left(\sum_{i=1}^{m} p_i \rho_i \right) (\sqrt{\rho})^{-1} \sqrt{\bar{\rho}_n} + \sqrt{\bar{\rho}_n} (\mathbf{I} - \mathbf{P}) \sqrt{\bar{\rho}_n} = \bar{\rho}_n. \end{aligned}$$

This proves the lemma. □

Proposition 12.1.4. *The functional* $\chi_\Phi(\cdot) : \mathcal{P}(\mathcal{S}(\mathbb{H}_A)) \to [0, +\infty]$ *is nonnegative, concave and lower semicontinuous such that*

$$\chi_\Phi(\bar{\rho}) - \sum_{i=1}^{+\infty} p_i \chi_\Phi(\rho_i) \geq \sum_{i=1}^{+\infty} p_i H(\bar{\rho}_i \| \Phi(\bar{\rho})) \tag{12.5}$$

for arbitrary ensemble $\mu = \{p_i, \rho_i\}_{i=1}^{+\infty}$ with the average state $\bar{\rho}(\mu) = \sum_{i=1}^{+\infty} p_i \rho_i$. Furthermore, if $H(\Phi(\overline{\rho}(\mu))) < +\infty$, then the following alternate expression for $\chi_\Phi(\mu)$ holds:

$$\chi_\Phi(\mu) = H(\Phi(\overline{\rho}(\mu))) - \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu(d\rho). \tag{12.6}$$

Proof. 1. The nonnegativity of the functional $\chi_\Phi(\cdot) : \mathcal{P}(\mathcal{S}(\mathbb{H}_A)) \to [0, +\infty]$ is trivial and the lower semicontinuity follows from the lower semicontinuity of the extended von Neumann entropy $H(\cdot) : \mathfrak{T}_+(\mathbb{H}_A) \to [0, +\infty]$.

Let us first show its concavity. Note that for a convex set of states with finite output entropy this concavity easily follows from (12.4). We now prove concavity on the whole state space $\mathcal{S}(\mathbb{H}_A)$ and the inequality (12.5) holds. Let $\epsilon > 0$ be arbitrary. By definition of the χ-function for each $i = 1, \ldots, n$, there exists ensemble $\{q_j^i, \sigma_j^i\}_{j=1}^{m(i)}$ with the average $\bar{\rho}_i = \sum_{j=1}^{m(i)} q_i \sigma_j$ such that $\chi_\Phi(\{q_j^i, \sigma_j^i\}) > \chi_\Phi(\rho_i) - \epsilon$. Since the average state of the ensemble $\sum_{i=1}^{n} p_i \{q_j^i, \sigma_j^i\}$ coincides with $\bar{\rho}$, we have

$$\begin{aligned}
\chi_\Phi(\bar{\rho}) &\geq \chi_\Phi\left(\sum_{i=1}^{n} p_i \{q_j^i, \sigma_j^i\}\right) \\
&\geq \sum_{i=1}^{n} p_i \chi_\Phi(\{q_j^i, \sigma_j^i\}) + \sum_{i=1}^{n} p_i H(\Phi(\bar{\rho}_i) \| \Phi(\bar{\rho})) \\
&\geq \sum_{i=1}^{n} p_i \chi_\Phi(\rho_i) + \sum_{i=1}^{n} p_i H(\Phi(\bar{\rho}_i) \| \Phi(\bar{\rho})) - \epsilon.
\end{aligned}$$

Since ϵ can be arbitrary small, inequality (12.5) is established. It obviously implies concavity of the χ-function.

To prove lower semicontinuity of the χ-function, we have to show that

$$\liminf_{n \to +\infty} \chi_\Phi(\rho_n) \geq \chi_\Phi(\rho_0), \tag{12.7}$$

for arbitrary state ρ_0 and arbitrary sequence $(\rho_n)_{n=1}^{+\infty}$ converging to this state ρ_0. For arbitrary $\epsilon > 0$, let $\{p_i, \rho_i\}$ be an ensemble with the average ρ_0 such that

$$\sum_i p_i H(\Phi(\rho_i) \| \Phi(\rho_0)) \geq \chi_\Phi(\rho_0) - \epsilon.$$

By Lemma 12.1.3, there exists the sequence of ensembles $\{p_i^n, \rho_i^n\}$ of fixed size such that

$$\lim_{n \to +\infty} p_i^n = p_i, \quad \lim_{n \to +\infty} \rho_i^n = \rho_i, \quad \text{and} \quad \rho_n = \sum_i p_i^n \rho_i^n.$$

By definition, we have

$$\liminf_{n\to+\infty}\chi_\Phi(\rho_n)\geq\liminf_{n\to+\infty}\sum_i p_i^n H(\Phi(\rho_i^n)\|\Phi(\rho_n))$$
$$\geq\sum_i p_i H(\Phi(\rho_i)\|\Phi(\rho_0))\geq\chi_\Phi(\rho_0)-\epsilon,$$

where lower semicontinuity of the relative entropy was used. This implies (12.7) (due to the freedom of the choice of ϵ).

2. If $H(\Phi(\overline{\rho}(\mu)))<+\infty$, let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be an arbitrary sequence of finite- dimensional projectors monotonously increasing to the unit operator $\mathbf{I}_B$ on the output Hilbert space $\mathbb{H}_B$. We first note that the mapping $\rho\mapsto H(\Phi(\rho)\|\Phi(\overline{\rho}(\mu)))$ is nonnegative and measurable. This is because

$$\rho\mapsto H\big(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n\|\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n\big)$$

is nonnegative and measurable for each n. Therefore, the mapping

$$\rho\mapsto H\big(\Phi(\rho)\|\Phi(\overline{\rho}(\mu))\big)=\lim_{n\to+\infty}H\big(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n\|\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n\big)$$

is nonnegative and measurable.

We next show that the functionals defined by

$$\chi_\Phi^{(n)}(\mu)=\int_{\mathcal{S}(\mathbb{H}_A)}H\big(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n\|\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n\big)\mu(d\rho),\quad n=1,2,\dots,$$

are continuous.

We first claim that

$$\mathrm{range}\big(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n\big)\subset\mathrm{range}\big(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n\big),$$

or equivalently

$$\big(\ker(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)\big)^\perp\subset\big(\ker(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)\big)^\perp,$$

for μ-almost all ρ. This is because for each n,

$$\begin{aligned}\phi\in\ker\big(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n\big)&\Rightarrow\big(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n\big)(\phi)=0\\&\Rightarrow\big(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n\big)(\phi)=0\text{ for }\mu\text{-almost all }\rho\\&\Rightarrow\phi\in\ker\big(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n\big)\text{ for }\mu\text{-almost all }\rho.\end{aligned}$$

This proves the claim that $(\ker(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n))^\perp\subset(\ker(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n))^\perp$ for μ-almost all ρ.

It follows from the definition of relative entropy $H(\cdot\|\cdot)$ of quantum states with finite rank (see Definition 8.1.3) that

$$\begin{aligned}
&H(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n\|\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)\\
&\quad=\operatorname{tr}[(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)\log(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)]-\operatorname{tr}[(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)\log(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)]\\
&\qquad+\operatorname{tr}[\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n]-\operatorname{tr}[\mathbf{P}_n\Phi(\rho)\mathbf{P}_n]
\end{aligned}$$

for μ-almost all ρ. By using the definition of the von Neumann entropy $H:\mathcal{S}(\mathbb{H}_A)\to[0,+\infty]$ (see Definition 7.1.1), we have

$$\begin{aligned}
\chi_\Phi^{(n)}(\rho)=&-\int_{\mathcal{S}(\mathbb{H}_A)}H(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)\mu(d\rho)\\
&+\int_{\mathcal{S}(\mathbb{H}_A)}\operatorname{tr}[\mathbf{P}_n\Phi(\rho)]\log(\operatorname{tr}[\mathbf{P}_n\Phi(\rho)])\mu(d\rho)\\
&-\int_{\mathcal{S}(\mathbb{H}_A)}\operatorname{tr}[\mathbf{P}_n\Phi(\rho)\mathbf{P}_n]\log(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)\mu(d\rho)\\
&+\int_{\mathcal{S}(\mathbb{H}_A)}\operatorname{tr}[\mathbf{P}_n\Phi(\overline{\rho}(\mu))]\mu(d\rho)-\int_{\mathcal{S}(\mathbb{H}_A)}\operatorname{tr}[\mathbf{P}_n\Phi(\rho)]\mu(d\rho).
\end{aligned}$$

It is easy to see that the last two terms cancel while the central term can be transformed into

$$\begin{aligned}
&-\int_{\mathcal{S}(\mathbb{H}_A)}\operatorname{tr}[(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)\log(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)]\mu(d\rho)\\
&\quad=\operatorname{tr}\Big[-\int_{\mathcal{S}(\mathbb{H}_A)}(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)\log(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)\mu(d\rho)\Big]\\
&\quad=H(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)-\operatorname{tr}[\mathbf{P}_n\Phi(\overline{\rho}(\mu))]\log(\operatorname{tr}[\mathbf{P}_n\Phi(\overline{\rho}(\mu))]).
\end{aligned}$$

Hence,

$$\begin{aligned}
\chi_\Phi^{(n)}(\mu)=&\,H(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)-\operatorname{tr}[\mathbf{P}_n\Phi(\overline{\rho}(\mu))]\log(\operatorname{tr}[\mathbf{P}_n\Phi(\overline{\rho}(\mu)])\\
&-\int_{\mathcal{S}(\mathbb{H}_A)}H(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)\mu(d\rho)+\int_{\mathcal{S}(\mathbb{H}_A)}\operatorname{tr}[\mathbf{P}_n\Phi(\rho)]\log(\operatorname{tr}[\mathbf{P}_n\Phi(\rho)])\mu(d\rho).
\end{aligned}$$

Continuity and boundedness of the von Neumann entropy in the finite-dimensional case and similar properties of the function $\rho\mapsto\operatorname{tr}[\mathbf{P}_n\Phi(\rho)]\log(\operatorname{tr}[\mathbf{P}_n\Phi(\rho)])$ implies the continuity of $\chi_\Phi^{(n)}(\mu)$.

By the monotone convergence theorem, the sequence of functionals $(\chi_\Phi^{(n)}(\mu))_{n=1}^{+\infty}$ is monotonously increasing and pointwise converges to $\chi_\Phi(\mu)$. Therefore, the functional $\chi_\Phi(\mu)$ is lower semicontinuous.

To prove equation (12.6), we use the facts that

$$\lim_{n\to+\infty}H(\mathbf{P}_n\Phi(\overline{\rho}(\mu))\mathbf{P}_n)=H(\Phi(\overline{\rho}(\mu)))$$

and

$$\lim_{n\to+\infty}\int_{\mathcal{S}(\mathbb{H}_A)} H(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n))\mu(d\rho) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu(d\rho)$$

due to monotone convergence theorem. For every ρ, the sequence $(\mathrm{tr}[\mathbf{P}_n\Phi(\rho)])_{n=1}^{+\infty}$ take values in $[0,1]$ and converges to $\mathrm{tr}[\Phi(\rho)] = 1$. Therefore,

$$\lim_{n\to+\infty}\mathrm{tr}[\mathbf{P}_n\Phi(\rho)]\log(\mathrm{tr}[\mathbf{P}_n\Phi(\rho)]) = 0.$$

In particular, the second term in (12.8) goes to 0. Since $|x\log x| < 1$ for all $x \in (0,1]$, the last term in (12.8) also goes to 0 by dominated convergence theorem. This shows that

$$\lim_{n\to+\infty}\chi_\Phi^{(n)}(\mu) = \chi_\Phi(\mu) = H(\Phi(\overline{\rho}(\mu))) - \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu(d\rho).$$

This proves the proposition. □

While (12.6) is valid under the condition $H(\Phi(\rho)) < +\infty$, we note that

$H(\Phi(\rho)) = +\infty$ does not imply $\chi_\Phi(\rho) = +\infty$, however. Indeed, it is easy to construct a channel Φ from a finite-dimensional system $\mathbb{H}_A$ into infinite- dimensional $\mathbb{H}_B$ such that $H(\Phi(\rho)) = +\infty$ for any $\rho \in \mathcal{S}(\mathbb{H}_A)$ as shown in the following example, due originally to Shirokov [141].

Example 12.1. Let $\Phi(\rho) = \frac{1}{2}\rho \oplus \frac{1}{2}\,\mathrm{tr}[\rho]\tau$, where τ is a fixed state with $H(\tau) = +\infty$. Then $H(\Phi(\rho)) = +\infty$. On the other hand, by the monotonicity property of the relative entropy (see Theorem 9.1.3)

$$\sum_i p_i H(\Phi(\rho_i)\|\Phi(\rho)) \le \sum_i p_i H(\rho_i\|\rho) \le \log(\dim(\mathbb{H}_A)) < +\infty$$

for arbitrary ensemble $\{p_i,\rho_i\}$, and hence, $\chi_\Phi(\rho) \le \log(\dim(\mathbb{H}_A)) < +\infty$ for any $\rho \in \mathcal{S}(\mathbb{H}_A)$.

For arbitrary state ρ such that $H(\Phi(\rho)) < +\infty$, the χ-function has the following representation:

$$\chi_\Phi(\rho) = H(\Phi(\rho)) - \hat{H}(\Phi(\rho)), \tag{12.8}$$

where

$$\hat{H}(\Phi(\rho)) = \inf_{\mu\in\mathcal{P}_{\{\rho\}}}\int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu(d\rho) = \inf_{\sum_i p_i\rho_i=\rho}\sum_i p_i H(\Phi(\rho_i)) \tag{12.9}$$

is a convex closure of the output entropy $H(\Phi(\rho))$.

In the finite-dimensional case, the output entropy function $H(\Phi(\cdot))$ and its convex closure (= convex hull) $\hat{H}(\Phi(\cdot))$ are continuous concave and convex functions on $\mathcal{S}(\mathbb{H}_A)$ correspondingly and the representation (12.8) is valid for all states. It follows in this case $\chi_\Phi(\cdot)$ is continuous and concave on $\mathcal{S}(\mathbb{H}_A)$. However, in the infinite-dimensional case the output entropy $H(\Phi(\cdot))$ is only lower semicontinuous, and hence, the function $\chi_\Phi(\cdot)$ is not continuous even in the case of the noiseless channel Φ, for which $\chi_\Phi(\cdot) = H(\Phi(\cdot))$. But it turns out that the function $\chi_\Phi(\cdot)$ for arbitrary channel Φ has properties similar to the properties of the output entropy function $H(\Phi(\cdot))$.

The conclusion of the next proposition follows easily from the definition of $\chi_\Phi(\cdot)$.

Proposition 12.1.5. *Let $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H}_A)$ be a closed subset. If $H(\Phi(\cdot)) : \mathcal{A} \to [0,+\infty]$ is continuous in $\|\cdot\|_1$-norm, then $\chi_\Phi(\cdot) : \mathcal{A} \to [0,+\infty]$ is also continuous in $\|\cdot\|_1$-norm.*

In the modern convex analysis (see, e. g., Rockafelar [131]), the notion of strong convexity (concavity) plays an essential role in this section. By using inequality $H(\rho\|\sigma) \geq \frac{1}{2}\|\rho - \sigma\|_1$ and Proposition 12.2.4, we obtain the following observation.

Corollary 12.1.6. *$\chi_\Phi(\cdot)$ is a strongly concave function on $\mathcal{S}(\mathbb{H}_A)$ in the following sense:*

$$\begin{aligned}\chi_\Phi(\lambda\rho_1 + (1-\lambda)\rho_2) \\ \geq \lambda\chi_\Phi(\rho_1) + (1-\lambda)\chi_\Phi(\rho_2) + \lambda(1-\lambda)\|\Phi(\bar{\rho}_1) - \Phi(\bar{\rho}_2)\|_1^2\end{aligned}$$

for arbitrary ρ_1 and ρ_2 in $\mathcal{S}(\mathbb{H})$.

The similarity of the properties of the functions $\chi_\Phi(\rho)$ and $H(\Phi(\rho))$ is stressed by the following analog of Simon's dominated convergence theorem (due originally to Simon [164]) for quantum entropy, which will be used later.

Corollary 12.1.7. *Let $(\rho_n)_{n=1}^{+\infty}$ be a sequence of states in $\mathcal{S}(\mathbb{H}_A)$, converging to the state ρ under the $\|\cdot\|_1$-norm and such that $\lambda_n\rho_n \leq \rho$ for some sequence $(\lambda_n)_{n=1}^{+\infty}$ of positive numbers, converging to* 1. *Then*

$$\lim_{n\to+\infty} \chi_\Phi(\rho_n) = \chi_\Phi(\rho).$$

Proof. The condition $\lambda_n\rho_n \leq \rho$ implies decomposition $\rho = \lambda_n\rho_n + (1-\lambda_n)\sigma_n$, where $\sigma_n = (1-\lambda_n)^{-1}(\rho - \lambda_n\rho_n)$ is a state. By concavity of the χ-function, we have

$$\begin{aligned}\chi_\Phi(\rho) &= \chi_\Phi(\lambda_n\rho_n + (1-\lambda_n)\sigma_n) \\ &\geq \lambda_n\chi_\Phi(\rho_n) + (1-\lambda_n)\chi_\Phi(\sigma_n) \geq \lambda_n\chi_\Phi(\sigma_n),\end{aligned}$$

which implies $\limsup_{n\to+\infty}\chi_\Phi(\rho_n) \leq \chi_\Phi(\rho)$. This and lower semicontinuity of the χ-function completes the proof. □

Example 12.2. Let $\mathbf{H}_B$ be a positive unbounded operator on the space $\mathbb{H}_B$ such that $\operatorname{tr}[\exp(-\beta\mathbf{H}_B)] < +\infty$ for all $\beta > 0$ and $h > 0$ be a positive number. Proposition 9.2.2 yields continuity of the restriction of the output entropy $H(\Phi(\cdot))$ to the subset $\mathcal{A} = \{\sigma \in \mathcal{S}(\mathbb{H}) \mid \operatorname{tr}[\Phi(\rho)\mathbf{H}_B] \leq h\}$.

12.2 Holevo χ-capacities

Let $\mathcal{A}$ be a certain closed subset of $\mathcal{S}(\mathbb{H}_A)$, and let Φ be a quantum channel from $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$ to system $\mathcal{S}(\mathbb{H}_B)$. In this section, we investigate properties of output Holevo quantity $\chi_\Phi(\mu)$, unconstrained channel χ-capacity $C_\chi(\Phi)$ and $\mathcal{A}$-constrained channel χ-capacity $C_\chi(\Phi; \mathcal{A})$ for various types of closed subsets $\mathcal{A}$ of $\mathcal{S}(\mathbb{H}_A)$. All of these channel χ-capacities are often referred to as Holevo χ-capacities (see Holevo [70]).

We first define the unconstrained channel χ-capacity $C_\chi(\Phi)$ and $\mathcal{A}$-constrained channel χ-capacity $C_\chi(\Phi; \mathcal{A})$ of Φ as follows.

Definition 12.2.1 (Unconstrained channel χ-capacity). Let $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ and let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a channel. We define the unconstrained Holevo χ-capacity $C_\chi(\Phi)$ of Φ as

$$C_\chi(\Phi) := C_\chi(\Phi; \mathcal{S}(\mathbb{H}_A)) = \sup_{\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))} \chi_\Phi(\mu). \tag{12.10}$$

Let $\mathcal{A}$ be a closed subset of $\mathcal{S}(\mathbb{H}_A)$. Recall from Section 3.3 that $\mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}_A))$ is a subset of $\mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ defined by

$$\mathcal{P}_\mathcal{A} := \mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}_A)) = \{\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A)) \mid \bar{\rho}(\mu) \in \mathcal{A}\}, \tag{12.11}$$

where $\bar{\rho}(\mu)$ is the barycenter of μ defined by

$$\bar{\rho}(\mu) := \int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu(d\rho) \in \mathcal{A}.$$

Definition 12.2.2 ($\mathcal{A}$-constrained channel). Let $\mathcal{A}$ be a closed subset of $\mathcal{S}(\mathbb{H}_A)$ such that (12.11) is satisfied and let $\Phi : \mathcal{A} \subset \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a channel constrained to $\mathcal{A}$. In this case, Φ is said to be an $\mathcal{A}$-constrained channel.

Definition 12.2.3 ($\mathcal{A}$-constrained channel χ-capacity). The χ-capacity of the $\mathcal{A}$-constrained channel Φ is defined as

$$C_\chi(\Phi; \mathcal{A}) = \sup_{\bar{\rho}(\mu) \in \mathcal{A}} \chi_\Phi(\mu), \tag{12.12}$$

where

$$\chi_\Phi(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho)\|\Phi(\bar{\rho}(\mu)))\mu(d\rho),$$

as defined in (12.4).

The capacities $C_\chi(\Phi)$ and $C_\chi(\Phi;\mathcal{A})$ defined above are also called unconstrained and constrained *Holevo χ-capacity*, respectively. It is obvious that $C_\chi(\Phi;\mathcal{A}) = C_\chi(\Phi)$ when $\mathcal{A} = \mathcal{S}(\mathbb{H}_A)$.

We shall only consider the constraint sets $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$ such that $C_\chi(\Phi;\mathcal{A}) < +\infty$ throughout this section.

The following result provides an equivalent definition of $C_\chi(\Phi;\mathcal{A})$.

Proposition 12.2.4. *The χ-capacity of $\mathcal{A}$-constrained channel Φ can be defined by*

$$C_\chi(\Phi;\mathcal{A}) = \sup_{\mu\in\mathcal{P}_\mathcal{A}} \chi_\Phi(\mu), \tag{12.13}$$

where $\mathcal{P}_\mathcal{A} = \mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}_A))$ is as defined (12.11).

Proof. We first note that

$$\sup_{\mu\in\mathcal{P}_\mathcal{A}} \chi_\Phi(\mu) \le C_\chi(\Phi;\mathcal{A}).$$

We now want to show that

$$\sup_{\mu\in\mathcal{P}_\mathcal{A}} \chi_\Phi(\mu) \ge C_\chi(\Phi;\mathcal{A}).$$

Suppose $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ is such that $\bar{\rho}(\mu) = \bar{\rho} \in \mathcal{A}$, i. e., $\mu \in \mathcal{P}_\mathcal{A}$. By Lemma 3.3.9, there exists a sequence of probability measures $(\mu_n)_{n=1}^{+\infty}$ on $\mathcal{S}(\mathbb{H}_A)$ with finite support and with the given barycenter $\bar{\rho}$ such that μ_n converges to μ under the weak convergence topology. Note that Definition 12.13 is a similar expression in which the supremum is over all measures in $\mathcal{P}_\mathcal{A}$ with finite support. By Proposition 12.1.4, the lower semicontinuity of $\chi_\Phi(\cdot)$ implies that $\lim_{n\to+\infty}\chi_\Phi(\mu_n) \ge \chi_\Phi(\mu)$. It follows that the supremum over all measures in $\mathcal{P}_\mathcal{A}$ coincides with the supremum over all measures in $\mathcal{P}_\mathcal{A}$ with finite support. Therefore,

$$\sup_{\mu\in\mathcal{P}_\mathcal{A}} \chi_\Phi(\mu) \ge C_\chi(\Phi;\mathcal{A}).$$

This proves the proposition. □

Based on the above proposition, the χ-capacity of the $\mathcal{A}$-constrained channel Φ can be defined as

$$C_\chi(\Phi;\mathcal{A}) = \sup_{\mu\in\mathcal{P}_\mathcal{A}} \chi_\Phi(\mu) = \sup_{\bar{\rho}(\mu)\in\mathcal{A}} \chi_\Phi(\mu), \tag{12.14}$$

where the short notation $\mathcal{P}_\mathcal{A} = \mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}_A))$ is used for simplicity.

Summarized from Definition 12.2.3 and Proposition 12.2.4, we have the following two observations:

1. When the probability measure μ has a countable (finite or infinite) support, then μ can be written as an ensemble $\mu = \{p_i, \rho_i\}_{i=1}^{+\infty}$, where $p_i = \mu(\{\rho_i\}) > 0$ and $\sum_{i=1}^{+\infty} p_i = 1$. If its barycenter (or average input state) $\bar{\rho}(\{p_i, \rho_i\}) = \sum_i p_i \rho_i \in \mathcal{A}$, then the χ-capacity of the $\mathcal{A}$-constrained channel Φ is defined as

$$C_\chi(\Phi; \mathcal{A}) = \sup_{\mu:\bar{\rho}(\{p_i,\rho_i\})\in\mathcal{A}} \chi_\Phi(\{p_i, \rho_i\}), \tag{12.15}$$

where

$$\chi_\Phi(\{p_i, \rho_i\}) = \sum_i p_i H\Big(\Phi(\rho_i) \| \Phi\Big(\sum_i p_i \rho_i\Big)\Big). \tag{12.16}$$

2. When $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ is a point-mass measure at $\rho \in \mathcal{S}(\mathbb{H}_A)$. This is, $\mu(\{\rho\}) = 1$, or equivalently $\mu = \{\rho\}$. Then $\bar{\rho}(\mu) = \rho$. This gives us the motivation to define

$$\chi_\Phi(\rho) := C_\chi(\Phi; \{\rho\}) = \sup_{\mu\in\mathcal{P}_{\{\rho\}}} \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\sigma)\|\Phi(\rho))\mu(d\sigma), \tag{12.17}$$

where $\mathcal{P}_{\{\rho\}}$ is the set of all probability measures on $\mathcal{S}(\mathbb{H})$ with the barycenter ρ, and that under the condition $H(\Phi(\rho)) < +\infty$ the supremum in (12.17) is achieved on some measure supported by pure states. The last assertion of this observation follows from the representation (12.8) and from lower semicontinuity of the function $\hat{H}_\Phi(\rho)$.

12.2.1 Optimal ensembles

For a closed subset $\mathcal{A}$ of $\mathcal{S}(\mathbb{H}_A)$, one of the questions that need to be answered is: under what condition(s) on $\mathcal{A}$, can the supremum in (12.14) be achieved? This subsection is devoted to answering this question.

To answer this question, we first introduce the concept of an optimal signal ensemble for the $\mathcal{A}$-constrained quantum channel as follows.

Definition 12.2.5. A probability measure $\mu^* \in \mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}_A))$ is said to be an optimal ensemble (or measure) for the $\mathcal{A}$-constrained channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ if the supremum in (12.14) is achieved by μ^*. That is,

$$C_\chi(\Phi; \mathcal{A}) = \chi_\Phi(\mu^*).$$

Definition 12.2.6. An optimal ensemble (or measure) $\mu^* \in \mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}_A))$ is said to be an optimal discrete ensemble for the $\mathcal{A}$-constrained channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ if

μ^* has a finite or countably infinite support $\{p_i, \rho_i\}$. In this case, the χ-capacity of the channel can be written as

$$\begin{aligned} C_\chi(\Phi; \mathcal{A}) = \chi_\Phi(\{p_i, \rho_i\}) &= \sum_i p_i H(\Phi(\rho_i) \| \Phi(\bar{\rho}(\mu^*))) \\ &= H(\Phi(\bar{\rho}(\mu^*))) - \sum_i p_i H(\Phi(\rho_i)) \\ &= H\Big(\Phi\Big(\sum_i p_i \rho_i\Big)\Big) - \sum_i p_i H(\Phi(\rho_i)). \end{aligned}$$

Definition 12.2.7. An optimal ensemble (or measure) $\mu^* \in \mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}_A))$ is said to be an optimal generalized ensemble for the $\mathcal{A}$-constrained channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ if μ^* does not have a finite or countably infinite support $\{p_i, \rho_i\}$. In this case, the χ-capacity of the channel can be expressed as

$$\begin{aligned} C_\chi(\Phi; \mathcal{A}) = \chi_\Phi(\mu^*) &= \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho) \| \Phi(\bar{\rho}(\mu^*)))\mu^*(d\rho) \\ &= H(\Phi(\bar{\rho}(\mu^*))) - \int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho))\mu^*(d\rho). \\ &= H\Big(\Phi\Big(\int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu^*(d\rho)\Big)\Big) - \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu^*(d\rho). \end{aligned}$$

We need Caratheodory's theorem (Caratheodory [17]) and Ulhman's result to prove Schumacher–Westmoreland theorem 12.2.9 below.

In convex geometry, Carathéodory's theorem roughly states (see Luo and Cao [110]) that if a point x of $\mathbb{R}^d$ lies in the convex hull of a set A, then x can be written as the convex combination of at most $d + 1$ points in A. Namely, there is a subset B of A consisting of $d+1$ or fewer points such that x lies in the convex hull of B. Equivalently, x lies in an r-simplex with vertices in A, where $r \le d$. The smallest r that makes the last statement valid for each x in the convex hull of A is defined as the Carathéodory's number of A. Depending on the properties of A, upper bounds lower than the one provided by Carathéodory's theorem can be obtained. Note that A need not be itself convex. A consequence of this is that B can always be extremal in A, as nonextremal points can be removed from A without changing the membership of x in the convex hull.

In the following, we formally state the Caratheodory theorem without a proof (see Caratheodory [17] and Steinitz [166] for a proof).

Theorem 12.2.8 (Caratheodory's theorem). *Suppose a point x lies in the convex hull of a set $A \subset \mathbb{R}^d$. There exists a subset $B \subset A$ containing no more than $d+1$ points such that x lies in the convex hull of B.*

The following existence of the optimal signal ensemble for finite-dimensional Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$ is originally due to Schumacher and Westmoreland [163] and Uhlmann [170].

Theorem 12.2.9 (Schumacher and Westmoreland [139], Uhlmann [170]). *Assume that* $\dim(\mathbb{H}_A) < +\infty$, $\dim(\mathbb{H}_B) < +\infty$, *and* $\mathcal{A}$ *is a convex compact subset of* $\mathcal{S}(\mathbb{H}_A)$. *Then an optimal signal ensemble exists.*

Proof. Assume that $\dim(\mathbb{H}_A) = d$ and let $\rho \in \mathcal{A}$. If the set of extremal elements of $\mathcal{A}$ is compact, then for any $\rho \in \mathcal{A}$, there exists an ensemble of states $\{\rho_k\} \subset \mathcal{A}$ with $\rho = \{p_k, \rho_k\}$ that maximizes $\chi_\Phi(\mu)$ over the set of all ensembles whose average state is $\bar{\rho}$. In other words, there exist optimal signal ensembles for a given average state $\bar{\rho}$. By Caratheodory's theorem 12.2.8, since the Hilbert space has d dimensions, then there are optimal ensembles (in this sense) with no more than d^2 states. We see that the conditions for the result from Uhlmann [170] are met. The set of states $\mathcal{A}$ that are possible outputs of the channel is a convex, compact set with a compact set of extremal points. For any average state $\bar{\rho}$ in $\mathcal{A}$, we can find a $\bar{\rho}$-fixed optimal ensemble with d^2 or fewer elements. Thus, in order to maximize $\chi_\Phi(\mu)$ over all possible ensembles, we only need to consider the set of ensembles with no more than d^2 elements drawn from $\mathcal{A}$. As this is a finite Cartesian product of a compact set, it is compact. As $\chi_\Phi(\cdot)$ is a continuous function, it must achieve its maximum in this set of ensembles. Thus, the existence of an optimal ensemble of states in $\mathcal{A}$ is assured. This proves the theorem. □

We define an approximating sequence of ensembles and its *optimal average state* for $C_\chi(\Phi; \mathcal{A})$ below.

Definition 12.2.10. The sequence of discrete ensembles $(\{p_i^k, \rho_i^k\}_{i=1}^{n(k)})_{k=1}^{+\infty}$ with the averages

$$\overline{\rho}^k := \sum_{i=1}^{n(k)} p_i^k \rho_i^k \in \mathcal{A}, \quad k = 1, 2, \ldots$$

is called an *approximating sequence* if $\lim_{k\to+\infty} \chi_\Phi(\{p_i^k, \rho_i^k\}) = C_\chi(\Phi; \mathcal{A})$. The state $\overline{\rho} \in \mathcal{A}$ is called an *optimal average state* if it is a partial limit of a sequence of average states $(\overline{\rho}^k)_{k=1}^{+\infty}$ for some approximating sequence of ensembles $(\{p_i^k, \rho_i^k\}_{i=1}^{n(k)})_{k=1}^{+\infty}$.

Note that if the set $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$ is compact, then the set of optimal average states is not empty. This is because $\mathcal{P}_\mathcal{A} = \mathcal{P}_\mathcal{A}(\mathcal{S}(\mathbb{H}_A))$ is a compact subset of $\mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ under weak convergence if and only if $\mathcal{A}$ is a compact subset of $\mathcal{S}(\mathbb{H}_A)$ under trace-norm $\|\cdot\|_1$ (see Proposition 3.2.8). Therefore, a sequence of discrete optimal (signal) ensemble $(\{p_i^k, \rho_i^k\}_{i=1}^{n(k)})_{k=1}^{+\infty}$ exists such that

$$C_\chi(\Phi; \mathcal{A}) = \lim_{k\to+\infty} \chi_\Phi(\{p_i^k, \rho_i^k\}_{i=1}^{n(k)}).$$

The following results are due originally to Holevo and Shirokov [79] (see also Holevo and Shirokov [81, 83]).

Theorem 12.2.11 ([79, 81, 83]). *If the restriction of the output entropy*

$$H(\Phi(\cdot)) : \mathcal{A} \subset \mathcal{S}(\mathbb{H}_A) \to [0, +\infty]$$

to the closed set $\mathcal{A}$ is continuous at least at one optimal average state $\overline{\rho}_0 \in \mathcal{A}$, then there exists an optimal generalized ensemble μ^* in $\mathcal{P}_{\mathcal{A}}$ such that $\operatorname{supp}(\mu^*) \subseteq \operatorname{extr}(\mathcal{S}(\mathbb{H}_A))$ *and*

$$C_\chi(\Phi; \mathcal{A}) = \chi_\Phi(\mu^*) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho) \| \Phi(\overline{\rho}(\mu^*)))\mu^*(d\rho).$$

Proof. (i) We first show that the mapping

$$\mu \mapsto \int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho))\mu(d\rho)$$

is well-defined and lower semicontinuous on the set $\mathcal{P}_{\mathcal{A}} := \mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}_A))$. Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be a sequence of increasing finite rank projection operators that converges to the identity operator $\mathbf{I}_B = \mathbf{I}_{\mathbb{H}_B}$ on the space $\mathbb{H}_B$. Then the output von Neumann entropy $H(\Phi(\rho))$ is a pointwise limit of the monotonously increasing sequence of functions, $f_n : \mathcal{S}(\mathbb{H}_A) \to [0, +\infty]$, defined by

$$f_n(\rho) = \operatorname{tr}[\eta(\mathbf{P}_n\Phi(\rho)\mathbf{P}_n)] - \eta(\operatorname{tr}[\mathbf{P}_n\Phi(\rho)\mathbf{P}_n]),$$

where

$$\eta(x) = \begin{cases} -x\log x & \text{for } x > 0 \\ 0 & \text{for } x = 0. \end{cases}$$

It is clear that the sequence of functions $(f_n(\cdot))_{n=1}^{+\infty}$ are continuous and bounded on $\mathcal{S}(\mathbb{H}_A)$. Hence, the function $H(\Phi(\rho))$ is measurable and the monotone convergence theorem implies that

$$\int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu(d\rho) = \lim_{n\to+\infty} \int_{\mathcal{S}(\mathbb{H}_A)} f_n(\rho)\mu(d\rho).$$

Since the sequence of continuous functionals

$$\mu \mapsto \int_{\mathcal{S}(\mathbb{H}_A)} f_n(\rho)\mu(d\rho)$$

is nondecreasing, its pointwise limit

$$\mu \mapsto \lim_{n\to+\infty} \int\limits_{\mathcal{S}(\mathbb{H}_A)} f_n(\rho)\mu(d\rho) = \int\limits_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu(d\rho)$$

is therefore well-defined and lower semicontinuous.

(ii) By the assumption of the theorem, the restriction of the function $H(\Phi(\cdot))$ to the set $\mathcal{A}$ is continuous at some optimal average state $\bar{\rho}_0$. Therefore, the continuity of the mapping $\mu \mapsto \bar{\rho}(\mu)$ implies that the restriction of the functional $\mu \mapsto H(\Phi(\bar{\rho}(\mu)))$ to the set $\mathcal{P}_{\mathcal{A}}$ is continuous at any point μ_0 such that $\bar{\rho}(\mu_0) = \bar{\rho}_0$. Consequently, $H(\Phi(\bar{\rho}(\mu))) < +\infty$ for any point μ in the intersection of $\mathcal{P}_{\mathcal{A}}$ with some neighborhood of μ_0. For every such point μ, the relation (12.6) holds. Therefore, the restriction of the functional $\chi_\Phi(\cdot)$ to the set $\mathcal{P}_{\mathcal{A}}$ is upper semicontinuous, and by Proposition 12.1.4 it is continuous at any point μ_0 in $\mathcal{P}_{\mathcal{A}}$ such that $\bar{\rho}(\mu_0) = \bar{\rho}_0$. Let $(\{p_i^n, \rho_i^n\})_{i=1}^{+\infty}$ be an approximating sequence of ensembles with the corresponding sequence of average states $\bar{\rho}^n$ converging to the state $\bar{\rho}_0$. Decomposing each state of the ensembles $(\{p_i^n, \rho_i^n\})_{i=1}^{+\infty}$ into a countable convex combinations of pure states, we obtain the sequence $(\{\hat{p}_j^n, \hat{\rho}_j^n\})_{j=1}^{+\infty}$ of generalized ensembles consisting of a countable number of pure states with the same sequence of the average states $\bar{\rho}^n$. Let $\hat{\mu}^n$ be the sequence of measures ascribing value $\hat{p}_j^n$ to the set $\{\hat{\rho}_j^n\}$ for each j. It follows that

$$\begin{aligned}\chi_\Phi(\hat{\mu}^n) &= \sum_j \hat{p}_j^n H(\Phi(\hat{\rho}_j^n)\|\Phi(\bar{\rho}^n)) \\ &\geq \sum_i p_i^n H(\Phi(\rho_i^n)\|\Phi(\bar{\rho}^n)) = \chi_\Phi(\{p_i^n, \rho_i^n\}),\end{aligned} \tag{12.18}$$

where the inequality follows from convexity of the relative entropy. By construction, $\operatorname{supp}(\hat{\mu}^n) \subseteq \operatorname{extr}(\mathcal{S}(\mathbb{H}))$ for each n. By Proposition 3.2.8, there exists a subsequence converging to some measure μ^{n_k} in $\mathcal{P}_{\mathcal{A}}$. Since the set $\operatorname{extr}(\mathcal{S}(\mathbb{H}))$ of all pure states is closed subset of $\mathcal{S}(\mathbb{H}_A)$, we have $\operatorname{supp}(\mu^*) \subseteq \operatorname{extr}(\mathcal{S}(\mathbb{H}))$. It is clear that $\bar{\rho}(\mu^*) = \bar{\rho}_0$, and hence, as shown above, the restriction of the functional $\chi_\Phi(\mu)$ on the set $\mathcal{P}_{\mathcal{A}}$ is continuous at the point μ^*. This and the approximating property of the sequence $(\{p_i^n, \rho_i^n\})$ and (12.18) implies

$$C_\chi(\Phi; \mathcal{A}) = \lim_{k\to\infty} \chi_\Phi(\{p_i^{n(k)}, \rho_i^{n(k)}\}) \leq \lim_{k\to+\infty} \chi_\Phi(\hat{\mu}_{n(k)}) = \chi_\Phi(\mu^*).$$

Since the converse inequality follows from Proposition 12.2.4, we obtain

$$C_\chi(\Phi; \mathcal{A}) = \chi_\Phi(\mu^*),$$

which means that the measure μ^* is an optimal generalized ensemble for the $\mathcal{A}$-constrained channel Φ. This proves the theorem. □

By noting that the condition of the theorem holds trivially for $\mathcal{A} = \{\rho_0\}$, we have the following corollary.

Corollary 12.2.12. *For arbitrary state ρ_0 with $H(\Phi(\rho_0)) < +\infty$, there exists a generalized ensemble μ_0 such that $\bar{\rho}(\mu_0) = \rho_0$ and*

$$\chi_\Phi(\{\rho_0\}) := \sup_{\sum_i p_i \rho_i = \rho_0} \chi_\Phi(\{p_i, \rho_i\}) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho) \| \Phi(\rho_0)) \mu_0(d\rho).$$

In the finite-dimensional case, we obviously have

$$C_\chi(\Phi; \mathcal{A}) = \chi_\Phi(\{\bar{\rho}\}), \tag{12.19}$$

where $\bar{\rho}$ is the average state of any optimal ensemble.

The generalization of this relation to the infinite-dimensional case is closely connected with the question of existence of the optimal generalized ensemble.

The first assertion of the next corollary is obvious while the second one follows from Corollary 12.2.12.

Corollary 12.2.13. *If an optimal generalized ensemble for the $\mathcal{A}$-constrained channel Φ exists, then the equality* (12.19) *holds for some optimal average state $\bar{\rho}$ for the A-constrained channel Φ. If the equality* (12.19) *holds for some optimal average state $\bar{\rho}$ for the $\mathcal{A}$-constrained channel Φ with $H(\Phi(\bar{\rho})) < +\infty$, then there exists an optimal generalized ensemble for the $\mathcal{A}$-constrained channel Φ.*

Remark 12.1. The continuity condition in Theorem 12.2.11 is essential, as is shown in Example 12.3 below. It is possible to show that this condition holds automatically if the set $\mathcal{A}$ is convex with a finite number of extreme points with finite output entropy. Shirokov [141] conjectured that this condition holds for an arbitrary convex compact set $\mathcal{A}$ due to the special properties of optimal average states in this case.

In the following, we consider existence and properties of optimal generalized ensembles.

The well-known fact concerning the χ-capacity of a finite-dimensional quantum channel Φ constrained by a closed subset $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$ consists in existence of an optimal ensemble at which the supremum in (12.14) is achieved. Since

$$C_\chi(\Phi; \mathcal{A}) = \sup_{\mu \in \mathcal{P}_\mathcal{A}} \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho) \| \Phi(\bar{\rho}(\mu))) \mu(d\rho), \tag{12.20}$$

where the supremum is taken over all probability measures μ with the barycenter $\bar{\rho}(\mu)$ in $\mathcal{A}$, the notion of an optimal ensemble is naturally generalized to the infinite-dimensional case leading to the notion of an optimal measure (generalized or continuous optimal ensemble) at which the supremum in (12.20) is achieved.

However, in contrast to the finite-dimensional case, we cannot claim existence of an optimal measure for an arbitrary quantum channel constrained by closed or even compact subset of $\mathcal{S}(\mathbb{H}_A)$ as demonstrated in the following example due originally to Holevo and Shirokov [79].

Example 12.3. Consider the Abelian von Neumann algebra l_∞ and its predual l_1. Let Φ be the noiseless channel on l_1. Consider the sequence of states $(\rho_n)_{n=1}^{+\infty}$,

$$\rho_n = \Big(1 - q_n, \underbrace{\frac{q_n}{n}, \frac{q_n}{n}, \ldots, \frac{q_n}{n}}_{n}, 0, 0, \ldots\Big), \quad n \in \mathbb{N},$$

where $(q_n)_{n=1}^{+\infty}$ is a sequence of numbers in $[0,1]$, which will be defined below. Note that in this case $\chi_\Phi(\rho_n) = H(\rho_n) = h_2(q_n) + q_n \log n$, where $h_2(x) = -x\log x - (1-x)\log(1-x)$. We will show later that there exists the sequence $(q_n)_{n=1}^{+\infty}$ such that $\lim_{n\to+\infty} q_n = 0$, while the corresponding sequence $\chi_\Phi(\rho_n) = H(\rho_n)$ monotonously increases to 1. Let q_n be such a sequence and $\mathcal{A}$ be the closure of the sequence ρ_n, which obviously consists of states ρ_n and pure state $\rho_* = \lim_{n\to+\infty} \rho_n = (1,0,0,\ldots)$. By definition and the above monotonicity $C_\chi(\Phi, \mathcal{A}) = \lim_{n\to+\infty} \chi_\Phi(\rho_n) = 1$, while ρ_* is the only optimal average state for the $\mathcal{A}$-constrained channel Φ and $\chi_\Phi(\rho_*) = H(\rho_*) = 0$. So, we have $C_\chi(\Phi; \mathcal{A}) > \chi_\Phi(\rho_*)$ and Corollary 12.2.13 implies that there is no optimal ensemble for the $\mathcal{A}$-constrained channel Φ. Let us construct the sequence q_n with the above properties. Consider the strongly increasing function $f(x) = x(1-\ln x)$ on $[0,1]$, where $\ln x$ denotes the natural logarithm function of x. It is easy to see that $f'(x) = -\ln x < 0$ for all $x \in [0,1]$ and $f([0,1]) = [0,1]$. Let f^{-1} be the inverse function of f and let $g(x) = xf^{-1}(\ln(2/x))$ for all $x \geq 1$. Note that the function $g(x)$ is implicitly defined by the equation

$$g\big(1 - \ln(g/x)\big) = \ln 2. \tag{12.21}$$

Using this, it is easy to see that the function $g(x)$ satisfies the following differential equation:

$$\ln(g/x)g' = g/x \tag{12.22}$$

Since $g(x)/x = f^{-1}(\ln 2/x)$, we have $g(x)/x \in [0,1]$. This with (12.21) and (12.22) implies $g(x) \in [0,1]$, $\lim_{x\to+\infty} g(x) = 0$ and $g'(x) < 0$, correspondingly. Consider the function $H(x) = h_2(g(x)) + g(x)\log x$. By (12.21) and (12.22) with the above observations, we have

$$\lim_{x\to+\infty} H(x) = (\ln 2)^{-1} \lim_{x\to+\infty} g(x)\ln x = 1$$

and

$$\begin{aligned} H'(x) &= (\ln 2)^{-1}\big(g'(x)\ln\big(1 - g(x)\big) - g'(x)\ln g(x) + g'(x)\ln x + g(x)/x\big) \\ &= (\ln 2)^{-1} g'(x)\log\big(1 - g(x)\big) > 0, \quad x > 1. \end{aligned}$$

It follows that $H(x)$ is an increasing function on $[1, +\infty[$, tending to its upper bound 1 at infinity. Setting $q_n = g(n)$, we obtain the sequence with the desired properties.

In the following subsections, we explore properties of the channel capacity $C_\chi(\Phi; \mathcal{A})$ with $\mathcal{A}$ being (i) a compact constraint subset and (ii) a convex constraint subset of $\mathcal{S}(\mathbb{H}_A)$.

12.2.2 Compact constraint $\mathcal{A}$

We first investigate properties of Holevo χ-quantity $\chi_\Phi(\mu)$ and Holevo χ-capacity $C_\chi(\Phi; \mathcal{A})$ when the input states belongs to compact constraint set $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$.

Recall that an unbounded positive operator $\mathbf{H}$ on $\mathbb{H}_A$ with discrete spectrum of finite multiplicity is called an $\mathfrak{H}$- operator. Let $\mathbf{Q}_n$ be the spectral projector of $\mathbf{H}$ corresponding to the lowest n eigenvalues of $\mathbf{H}$. We define

$$\operatorname{tr}[\rho \mathbf{H}] = \lim_{n \to +\infty} \operatorname{tr}[\rho \mathbf{Q}_n \mathbf{H}], \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A).$$

It has been shown in Theorem 3.2.5 that there exists an $\mathfrak{H}$-operator $\mathbf{H}$ and a positive constant h such that the following set:

$$\mathcal{K}_{\mathbf{H}}(h) = \{\rho \in \mathcal{S}(\mathbb{H}_A) \mid \operatorname{tr}[\rho \mathbf{H}] \le h\}$$

is a compact subset of $\mathcal{S}(\mathbb{H}_A)$ and $\mathcal{K}_{\mathbf{H}}(h) = \mathcal{A}$.

Proposition 12.2.14. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on the space* $\mathbb{H}_B$ *such that*

$$\operatorname{tr}[\exp(-\beta \mathbf{H})] < +\infty, \quad \beta > 0, \tag{12.23}$$

and $\operatorname{tr}[\Phi(\rho)\mathbf{H}] \le h$ *for all* $\rho \in \mathcal{A}$*. Then there exists an optimal generalized ensemble for the* $\mathcal{A}$*-constrained channel* Φ*.*

Proof. To show the existence of an optimal generalized ensemble for the $\mathcal{A}$-constrained channel Φ, we will show that under the condition of the proposition the restriction of the output entropy $H(\Phi(\cdot))$ on the set $\mathcal{A}$ is continuous. By applying Theorem 12.2.11, the conclusion of the proposition follows.

Let $\rho_\beta = (\operatorname{tr}[\exp(-\beta \mathbf{H})])^{-1} \exp(-\beta \mathbf{H})$ be a state in $\mathcal{S}(\mathbb{H}_B)$. For arbitrary ρ in $\mathcal{A}$, we have

$$H(\Phi(\rho) \| \rho_\beta) = -H(\Phi(\rho)) + \beta \operatorname{tr}[\Phi(\rho)\mathbf{H}] + \log(\operatorname{tr}[\exp(-\beta \mathbf{H})]). \tag{12.24}$$

Let ρ_n be an arbitrary sequence of states in $\mathcal{A}$ converging to the state ρ. By using (12.24) and lower semicontinuity of the relative entropy, we obtain

$$\begin{aligned}\limsup_{n\to+\infty} H(\Phi(\rho_n)) &= H(\Phi(\rho)) + H(\Phi(\rho)\|\rho_\beta) - \liminf_{n\to+\infty} H(\Phi(\rho_n)\|\rho_\beta)\\ &\quad + \limsup_{n\to+\infty} \beta\operatorname{tr}[\Phi(\rho_n)\mathbf{H}] - \beta\operatorname{tr}[\Phi(\rho)\mathbf{H}]\\ &\leq H(\Phi(\rho)) + \beta h.\end{aligned}$$

By letting $\beta \to 0$ in the above inequality, we can establish the upper semicontinuity of the restriction of the function $H(\Phi(\cdot))$ to the set $\mathcal{A}$. The lower semicontinuity of this function follows from the lower semicontinuity of the entropy. Hence, the restriction of the function $H(\Phi(\cdot))$ on the set $\mathcal{A}$ is continuous. This proves the proposition. □

12.2.3 Convex constraint $\mathcal{A}$

We study in this subsection properties of $\chi_\Phi(\mu)$ and $C_\chi(\Phi;\mathcal{A})$ when $\mathcal{A}$ is a convex subset of $\mathcal{S}(\mathbb{H}_A)$.

The following result is due originally to Donald [40]. The proof will be omitted, because a generalization of Donald's identity will be stated and proved following the lemma.

Lemma 12.2.15 (Donald identity). *Assume that* $\dim(\mathbb{H}_A) < \infty$ *and let* $\mu = \{p_i, \rho_i\}$ *be a discrete ensemble. Then the following identity holds for arbitrary* $\sigma \in \mathcal{S}(\mathbb{H}_A)$*:*

$$\sum_i p_i H(\rho_i\|\sigma) = \sum_i p_i H(\rho_i\|\overline{\rho}(\mu)) + H(\overline{\rho}(\mu)\|\sigma),$$

where $\overline{\rho}(\mu) = \sum_i p_i\rho_i$.

The following lemma is a generalization of Donald's identity lemma 12.2.15.

Lemma 12.2.16. *For arbitrary measure* $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ *and arbitrary state* $\sigma \in \mathcal{S}(\mathbb{H}_A)$*, the following identity holds:*

$$\int_{\mathcal{S}(\mathbb{H}_A)} H(\rho\|\sigma)\mu(d\rho) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\rho\|\overline{\rho}(\mu))\mu(d\rho) + H(\overline{\rho}(\mu)\|\sigma). \tag{12.25}$$

Proof. We first notice that in the finite-dimensional case Donald's identity

$$\sum_i p_i H(\rho_i\|\sigma) = \sum_i p_i H(\rho_i\|\overline{\rho}(\mu)) + H(\overline{\rho}(\mu)\|\sigma)$$

holds for not necessarily normalized positive operators with the generalized definition of the relative entropy. This can be obviously extended to generalized ensembles in a finite-dimensional Hilbert space, giving (12.25) for this case. Thus, this relation holds for the operators $\mathbf{P}_n\rho\mathbf{P}_n$, $\mathbf{P}_n\sigma\mathbf{P}_n$, where $(\mathbf{P}_n)_{n=1}^{+\infty}$ is an arbitrary sequence of finite projectors increasing to $\mathbf{I}_A = \mathbf{I}_{\mathbb{H}_A}$. Passing to the limit $n \to +\infty$ and using the monotone convergence theorem, we obtain (12.25) in the infinite-dimensional case. This proves the lemma. □

The following proposition is a generalization of the "maximal distance property" (see, e. g., Proposition 1 in Holevo and Shirokov [79]).

Proposition 12.2.17. *Let $\mathcal{A}$ be a closed convex subset of $\mathcal{S}(\mathbb{H}_A)$. A measure $\pi \in \mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}_A))$ is an optimal generalized ensemble for the $\mathcal{A}$-constrained channel Φ if and only if*

$$\int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho)\|\Phi(\bar{\rho}(\mu)))\mu(d\rho) \leq \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho)\|\Phi(\bar{\rho}(\pi)))\pi(d\rho) = \chi_\Phi(\mu) \tag{12.26}$$

for arbitrary measure $\mu \in \mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}_A))$.

Proof. ($\Rightarrow$) Suppose inequality (12.26) holds for arbitrary measure $\mu \in \mathcal{P}_{\mathcal{A}} = \mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}))$. By Lemma 12.2.16, we have

$$\begin{aligned}\chi_\Phi(\mu) &= \int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho)\|\Phi(\bar{\rho}(\mu)))\mu(d\rho) \\ &\leq \int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho)\|\Phi(\bar{\rho}(\mu)))\mu(d\rho) + H(\Phi(\bar{\rho}(\mu))\|\Phi(\bar{\rho}(\pi))) \\ &= \int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho)\|\Phi(\bar{\rho}(\pi)))\mu(d\rho) \leq \chi_\Phi(\pi),\end{aligned}$$

which implies optimality of the measure π.

($\Leftarrow$) Conversely, let π be an optimal generalized ensemble for the $\mathcal{A}$-constrained channel Φ and μ be an arbitrary measure in $\mathcal{P}_{\mathcal{A}}$. By convexity of the set $\mathcal{A}$, the measure $\pi_\lambda = \lambda\mu + (1-\lambda)\pi$ is also in $\mathcal{P}_{\mathcal{A}}$ for arbitrary $\lambda \in (0,1)$. Using Lemma 12.2.16, we have

$$\begin{aligned}\chi_\Phi(\pi_\lambda) &= \int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho)\|\Phi(\bar{\rho}(\pi_\lambda)))\pi_\lambda(d\rho) \\ &= \lambda \int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho)\|\Phi(\bar{\rho}(\pi_\lambda)))\mu(d\rho) \\ &\quad + (1-\lambda)\chi_\Phi(\pi) + (1-\lambda)H(\bar{\rho}(\pi)\|\bar{\rho}(\pi_\lambda)).\end{aligned}$$

The optimality of π and nonnegativity of the relative entropy imply

$$\int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho)\|\Phi(\bar{\rho}(\pi_\lambda)))\mu(d\rho) - \chi_\Phi(\pi) \leq \lambda^{-1}(\chi_\Phi(\pi_\lambda) - \chi_\Phi(\pi)) \leq 0. \tag{12.27}$$

By Lemma 12.2.16 and lower semicontinuity of relative entropy,

$$\begin{aligned}
&\liminf_{\lambda\to 0} \int_{\mathcal{S}(\mathbb{H})} H\big(\Phi(\rho)\|\Phi(\overline{\rho}(\pi_\lambda))\big)\mu(d\rho) \\
&= \int_{\mathcal{S}(\mathbb{H})} H\big(\Phi(\rho)\|\Phi(\overline{\rho}(\mu))\big)\mu(d\rho) + \liminf_{\lambda\to 0} H\big(\Phi(\overline{\rho}(\mu))\|\Phi(\overline{\rho}(\pi_\lambda))\big) \\
&\geq \int_{\mathcal{S}(\mathbb{H})} H\big(\Phi(\rho)\|\Phi(\overline{\rho}(\mu))\big)\mu(d\rho) + H\big(\Phi(\overline{\rho}(\mu))\|\overline{\rho}(\pi)\big) \\
&= \int_{\mathcal{S}(\mathbb{H})} H\big(\Phi(\rho)\|\Phi(\overline{\rho}(\pi))\big)\mu(d\rho).
\end{aligned}$$

Then (12.27) implies

$$\begin{aligned}
&\int_{\mathcal{S}(\mathbb{H})} H\big(\Phi(\rho)\|\Phi(\overline{\rho}(\pi))\big)\mu(d\rho) - \chi_\Phi(\pi) \\
&\quad\leq \liminf_{\lambda\to 0} \int_{\mathcal{S}(\mathbb{H})} H\big(\Phi(\rho)\|\Phi(\overline{\rho}(\pi_\lambda))\big)\mu(d\rho) - \chi_\Phi(\pi) \\
&\quad\leq \liminf_{\lambda\to 0}\big(\chi_\Phi(\pi_\lambda) - \chi_\Phi(\pi)\big) \leq 0.
\end{aligned}$$

This proves the proposition. □

Denote by $\mathcal{P}_p(\mathcal{S}(\mathbb{H}_A))$ the set of Borel probability measures on $\mathcal{S}(\mathbb{H}_A)$ supported by the set of pure states in $\mathcal{S}(\mathbb{H}_A)$.

By monotonicity of the relative entropy $H(\cdot\|\cdot)$ for an arbitrary quantum channel $\Phi : \mathfrak{T}(\mathbb{H}_A) \to \mathfrak{T}(\mathbb{H}_B)$, we have

$$\chi(\mu \circ \Phi) \leq \chi(\mu). \tag{12.28}$$

Note that if $H(\bar{\rho}) < +\infty$ and $H(\Phi(\bar{\rho})) < +\infty$, then inequality (12.28) means convexity of the entropy gain $H(\Phi(\rho)) - H(\rho)$ of the channel Φ. A necessary condition for the equality in (12.28) expressed in terms of the channel Φ is originally due to Petz [126]. However, the von Neumann entropy, despite the fact that it is more often called quantum entropy, is really a measure of the mixing property in a quantum mixed state, and thus measures the classicality of a quantum state. Indeed, the von Neumann entropy of any pure quantum state is zero.

We will use the following lemmas concerning the limits of von Neumann entropy $H(\cdot)$ and conditional entropy $H(\cdot\|\cdot)$. In what follows, convergence of quantum states means convergence of the corresponding density operators to a limit operator in the trace norm, which is equivalent to a weak operator convergence (cf. Davis [30] and Holevo and Shrikov [81]). Note that entropy and conditional entropy are lower semicontinuous functions (see Wehrl [175]).

Lemma 12.2.18. *Let $\mathcal{A} \subset \mathcal{S}(\mathbb{H}_A)$ be a set such that $C_\chi(\Phi;\mathcal{A}) < +\infty$ and ρ_B be a state in $\mathcal{S}(\mathbb{H}_B)$ such that*

$$\sum_j q_j H(\Phi(\sigma_j)\|\rho_B) \le C_\chi(\Phi;\mathcal{A})$$

for arbitrary ensemble $\nu = \{q_j,\sigma_j\}$ with the average $\bar{\sigma}(\nu) \in \mathcal{A}$. Then for arbitrary approximating sequence $(\{p_i^k,\rho_i^k\})_{k=1}^{+\infty}$ of ensembles for the $\mathcal{A}$-constrained channel Φ with the corresponding sequence of average states $\bar{\rho}^k$ there exists $\lim_{k\to+\infty} H(\Phi(\bar{\rho}^k)\|\rho_B) = 0$.

Proof. Let $(\{p_i^k,\rho_i^k\})_{k=1}^{+\infty}$ an approximating sequence of ensembles with the corresponding sequence of the average states $\bar{\rho}^k$. By assumption, we have

$$\sum_i p_i^k H(\Phi(\rho_i^K)\|\rho_b) \le C_\chi(\Phi;\mathcal{A}).$$

Applying Donald's identity (12.25) to the left-hand side, we obtain

$$\sum_i p_i^k H(\Phi(\rho_i^k)\|\rho_B) = \sum_i p_i^k H(\Phi(\rho_i^k)\|\Phi(\bar{\rho}^k)) + H(\Phi(\bar{\rho}^k)\|\rho_B) \tag{12.29}$$

From the above two expressions, we have

$$H(\Phi(\bar{\rho}^k)\|\rho_B) \le C_\chi(\Phi;\mathcal{A}) - \sum_i p_i H(\Phi(\rho_i^k)\|\Phi(\bar{\rho}^k)).$$

But the right-hand side tends to zero as k tends to infinity due to the approximating property of the sequence $(\{p_i^k,\rho_i^k\})_{k=1}^{+\infty}$. This proves the lemma. □

Despite possible nonexistence of partial limits of the sequence of the average states of a particular approximating sequence, the following proposition guarantees convergence of the sequence of their images.

Proposition 12.2.19. *Let $\mathcal{A}$ be convex subset of $\mathcal{S}(\mathbb{H})$ such that $C_\chi(\Phi;\mathcal{A}) < +\infty$. Then there exists a unique state $\Omega(\Phi,\mathcal{A})$ in $\mathcal{S}(\mathbb{H}_B)$ such that*

$$\sup_{\sum_j q_j\sigma_j\in\mathcal{A}} \sum_j q_j H\big(\Phi(\sigma_j)\|\Omega(\Phi,\mathcal{A})\big) = C_\chi(\Phi;\mathcal{A}),$$

where the supremum is taken over all discrete ensembles $\{(q_j,\sigma_j)\}$ with barycenter (or average state) $\sum_j q_j\sigma_j \in \mathcal{A}$. For arbitrary approximating sequence of ensembles $(\{p_i^k,\rho_i^k\}_{i=1}^{n(k)})_{k=1}^{+\infty}$ for the $\mathcal{A}$-constrained channel Φ, the following limit exists:

$$\lim_{k\to+\infty} \Phi(\bar{\rho}^k) = \lim_{k\to+\infty} \Phi\left(\sum_{i=1}^{n(k)} p_i^k\rho_i^k\right) = \Omega(\Phi,\mathcal{A}).$$

Proof. We first show that, for arbitrary approximating sequence of ensembles $(\mu^k)_{k=1}^{+\infty} = (\{p_i^k, \rho_i^k\}_{i=1}^{n(k)})_{k=1}^{+\infty}$ for the $\mathcal{A}$-constrained channel Φ, the sequence $(\Phi(\bar{\rho}^k))_{k=1}^{+\infty}$ converges to a particular state in $\mathcal{S}(\mathbb{H}_B)$, where $\bar{\rho}^k = \sum_{i=1}^{n(k)} p_i^k \rho_i^k$ for $k = 1, 2, \ldots$. By definition of an approximating sequence, for arbitrary $\epsilon > 0$ there exists a positive integer N_ϵ such that $\chi_\Phi(\mu^k) > C_\chi(\Phi; \mathcal{A}) - \epsilon$ for all $k \geq N_\epsilon$. By Lemma 12.2.15 with $m = 2$ and $\lambda = 1/2$ for all $k_1, k_2 \geq N_\epsilon$, we have

$$\begin{aligned} C_\chi(\Phi; \mathcal{A}) - \epsilon &\leq \frac{1}{2}\chi_\Phi(\mu^{k_1}) + \frac{1}{2}\chi_\Phi(\mu^{k_2}) \\ &\leq \chi_\Phi\left(\frac{1}{2}\mu^{k_1} + \frac{1}{2}\mu^{k_2}\right) - \frac{1}{8}\|\Phi(\bar{\rho}^{k_1}) - \Phi(\bar{\rho}^{k_2})\|_1 \\ &\leq C_\chi(\Phi, \mathcal{A}) - \frac{1}{8}\|\Phi(\bar{\rho}^{k_1}) - \Phi(\bar{\rho}^{k_1})\|_1, \end{aligned}$$

and hence, $\|\Phi(\bar{\rho}^{k_1}) - \Phi(\bar{\rho}^{k_2})\|_1 < \sqrt{8}\epsilon$. Thus, the sequence $(\Phi(\bar{\rho}^k))_{k=1}^{+\infty}$ is a Cauchy sequence, and hence, it converges to a particular state ρ_B in $\mathcal{S}(\mathbb{H}_B)$. Let $\nu = \{q_j, \sigma_j\}_{j=1}^m$ be an arbitrary ensemble with the average $\bar{\sigma}(\nu) \in \mathcal{A}$. Consider the family of ensembles

$$\mu_k^\lambda = (1-\lambda)\{p_i^k, \rho_i^k\}_{i=1}^{n(k)} + \lambda\{q_j, \sigma_j\}_{j=1}^m, \quad \forall \lambda \in [0,1] \text{ and } \forall k \in \mathbb{N} \tag{12.30}$$

with the average states $\bar{\rho}_k^\lambda$. By convexity of $\mathcal{A}$, we have

$$\bar{\rho}_k^\lambda = (1-\lambda)\sum_{i=1}^{n(k)} p_i^k \rho_i^k + \lambda \sum_{j=1}^m q_j^k \sigma_j^k \in \mathcal{A}, \quad \forall \lambda \in [0,1] \text{ and } \forall k \in \mathbb{N}.$$

By the above observation,

$$\begin{aligned} \lim_{k\to+\infty} \Phi(\bar{\rho}_k^\lambda) &= \lim_{k\to+\infty} \Phi\left((1-\lambda)\sum_{i=1}^{n(k)} p_i^k \rho_i^k + \lambda \sum_{j=1}^m q_j^k \sigma_j^k\right) \\ &= (1-\lambda)\lim_{k\to+\infty} \Phi\left(\sum_{i=1}^{n(k)} p_i^k \rho_i^k\right) + \lambda\Phi\left(\sum_{j=1}^m q_j^m \sigma_j^k\right) \\ &= (1-\lambda)\rho_B + \lambda\Phi(\bar{\sigma}). \end{aligned} \tag{12.31}$$

By definition,

$$\chi_\Phi(\mu_k^\lambda) = (1-\lambda)\sum_{i=1}^{n(k)} p_i^k H(\Phi(\rho_i^k)\|\Phi(\bar{\rho}_k^\lambda)) + \lambda \sum_{j=1}^m q_j H(\Phi(\sigma_j)\|\Phi(\bar{\rho}_k^\lambda)). \tag{12.32}$$

Since $C_\chi(\Phi; \mathcal{A}) < +\infty$, both sums on the right-hand side of the above expression are finite. Applying Donald's identity (Lemma 12.2.15) to the first sum, we obtain

$$\sum_{i=1}^{n(k)} p_i^k H(\Phi(\rho_i^k)\|\Phi(\bar{\rho}_k^\lambda)) = \chi_\Phi(\mu_k^0) + H(\Phi(\bar{\rho}_k)\|\Phi(\bar{\rho}_k^\lambda)).$$

Substitution of the above expression into (12.32) gives

$$\chi_\Phi(\mu_k^\lambda) = \chi_\Phi(\mu_k^0) + (1-\lambda)H(\Phi(\bar\rho_k)\|\Phi(\bar\rho_k^\lambda)) + \lambda\left(\sum_{j=1}^m q_j H(\Phi(\sigma_j)\|\Phi(\bar\rho_k^\lambda)) - \chi_\Phi(\mu_k^0)\right).$$

Due to nonnegativity of the relative entropy, it follows that for $\lambda \neq 0$,

$$\sum_{j=1}^m q_j H(\Phi(\sigma_j)\|\Phi(\bar\rho_k^\lambda)) \le \lambda^{-1}(\chi_\Phi(\mu_k^\lambda) - \chi_\Phi(\mu_k^0)) + \chi_\Phi(\mu_k^0). \tag{12.33}$$

By definition of the approximating sequence, we have

$$\lim_{k\to+\infty} \chi_\Phi(\mu_k^0) = C_\chi(\Phi;\mathcal{A}) \ge \chi_\Phi(\mu_k^\lambda), \quad \forall k. \tag{12.34}$$

It follows that

$$\liminf_{\lambda\downarrow 0} \liminf_{k\to+\infty} \lambda^{-1}(\chi_\Phi(\mu_k^\lambda) - \chi_\Phi(\mu_k^0)) \le 0. \tag{12.35}$$

By lower semicontinuity of the relative entropy, equations (12.32), (12.33), (12.34) and (12.35) imply

$$\sum_{j=1}^m q_j H(\Phi(\sigma_j)\|\rho_B) \le \liminf_{\lambda\downarrow 0} \liminf_{k\to+\infty} H(\Phi(\sigma_j)\|\Phi(\bar\rho_k^\lambda)) \le C_\chi(\Phi;\mathcal{A}).$$

This proves that

$$\sup_{\sum_{j=1}^m q_j\sigma_j\in\mathcal{A}} \sum_j H(\Phi(\sigma_j)\|\rho_B) \le C_\chi(\Phi;\mathcal{A}). \tag{12.36}$$

To prove the converse inequality, we consider an approximating sequence $(\{p_i^k,\rho_i^k\})_{k=1}^{+\infty}$. Applying Donald's identity 12.2.15, we obtain

$$\sum_i p_i^k H(\Phi(\rho_i^k)\|\rho_B) = \sum_i p_i^k H(\Phi(\rho_i^k)\|\Phi(\bar\rho^k)) + H(\Phi(\bar\rho^k)\|\rho_B).$$

By the approximating property of the sequence $(\{p_i^k,\rho_i^k\})_{k=1}^{+\infty}$, the first term on the right-hand side tends to $C_\chi(\Phi;\mathcal{A})$ as $k \to +\infty$, while the second is nonnegative. This proves "≥," and hence, "=" in (12.36).

By inequality $H(\rho\|\sigma) \ge \frac{1}{2}\|\rho - \sigma\|_1$ and Lemma 12.2.18, inequality (12.36) implies that, for arbitrary approximating sequence of ensembles $(\{q_j^k,\sigma_j^k\})_{k=1}^{+\infty}$ for the $\mathcal{A}$-constrained channel Φ, the corresponding sequence $\Phi(\bar\sigma^k)$ converges to the state ρ_B. Thus, this state ρ_B does not depend on the choice of an approximating sequence, so it is determined only by the channel Φ and by the constraint set $\mathcal{A}$. Denote this state by $\Omega(\Phi, \mathcal{A})$. Lemma 12.2.18 implies also that $\rho_B = \Omega(\Phi, \mathcal{A})$ is the unique state for which equality in (12.36) holds. This proves the proposition. □

Proposition 12.2.19 shows in particular that if the set of input average states for the $\mathcal{A}$-constrained channel Φ is nonempty, then it maps by the channel Φ into a single state.

Corollary 12.2.20. *If there exists input optimal average state $\bar{\rho}$ for the $\mathcal{A}$- constrained channel Φ, then $\Phi(\bar{\rho}) = \Omega(\Phi, \mathcal{A})$.*

Note that compactness of the set $\mathcal{A}$ guarantees existence of at least one input average state.

This corollary justifies the following definition.

Definition 12.2.21. The state $\Omega(\Phi, \mathcal{A})$ is called the output optimal average state for the $\mathcal{A}$-constrained channel Φ.

There exist examples of constrained channels with finite χ-capacity but with no input optimal average state, for which the output optimal average state is explicitly determined and plays an important role in studying of this channels.

Corollary 12.2.22. *Let $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H}_A)$ be a convex set. Then*

$$C_\chi(\Phi; \mathcal{A}) \geq \chi_\Phi(\rho) + H\big(\Phi(\rho)\|\Omega(\Phi; \mathcal{A})\big), \quad \forall \rho \in \mathcal{A}.$$

Proof. It is sufficient to consider the case $C_\chi(\Phi; \mathcal{A}) < +\infty$. Let $\{(p_i, \rho_i)\}$ be an arbitrary ensemble such that $\sum_i p_i\rho_i = \bar{\rho} \in \mathcal{A}$. By Proposition 12.2.19,

$$\sum_i p_i H\big(\Phi(\rho_i)\|\Omega(\Phi, \mathcal{A})\big) \leq C_\chi(\Phi; \mathcal{A}).$$

This inequality and Donald's identity (Lemma 12.2.15), we have

$$\sum_i p_i H\big(\Phi(\rho_i)\|\Omega(\Phi; \mathcal{A})\big) = \chi_\Phi\big(\{p_i, \rho_i\}\big) + H\big(\Phi(\rho)\|\Omega(\Phi, \mathcal{A})\big).$$

This completes the proof. □

There exists another approach to the definition of the state $\Omega(\Phi, \mathcal{A})$. It is possible to show that finiteness of $C_\chi(\Phi; \mathcal{A})$ implies the compactness of the set $\overline{\Phi(\mathcal{A})}$, where $\overline{\Phi(\mathcal{A})}$ is the closure of $\Phi(\mathcal{A}) \subseteq \mathcal{S}(\mathbb{H}_B)$ under the $\|\cdot\|_1$-norm. For the arbitrary ensemble $\nu = \{(q_j, \sigma_j)\}$ with the average/barycenter $\bar{\sigma}(\nu) = \sum_j q_j\sigma_j \in \mathcal{A}$, consider the lower semicontinuous function $F_{\{q_j,\sigma_j\}}(\cdot) = \sum_j q_j H(\sigma_j\|\cdot)$ on the set $\overline{\Phi(\mathcal{A})}$. The function $F(\cdot) = \sup_{\sum_j q_j\sigma_j \in \mathcal{A}} F_{\{q_j,\sigma_j\}}(\cdot)$ is also lower semicontinuous function on the set $\overline{\Phi(\mathcal{A})}$, and hence, the minimum on this set is achieved.

The following proposition asserts, in particular, the state $\Omega(\Phi, \mathcal{A})$ can be defined as the unique minimal point of the function $F(\cdot)$ defined in the above paragraph.

Proposition 12.2.23. *Let $\mathcal{A} \subseteq \mathcal{S}(\mathbb{H}_B)$ be a convex set. The χ-capacity of the $\mathcal{A}$-constrained channel Φ can be expressed as*

$$C_\chi(\Phi; \mathcal{A}) = \inf_{\rho_B \in \overline{\Phi(\mathcal{A})}} \Big(\sup_{\sum_j q_j \sigma_j \in \mathcal{A}} \sum_j q_j H(\Phi(\sigma_j) \| \rho_B) \Big).$$

If $C_\chi(\Phi; \mathcal{A}) < +\infty$, then $\Omega(\Phi, \mathcal{A})$ is the only state on which the infimum on the right-hand side is achieved.

Proof. If $C_\chi(\Phi; \mathcal{A}) < +\infty$, then $F(\Omega(\Phi, \mathcal{A})) = C_\chi(\Phi; \mathcal{A})$ due to Proposition 12.2.19. Let ρ_B be a state such that

$$\sup_{\sum_j q_j \sigma_j \in \mathcal{A}} \sum_j q_j H(\Phi(\sigma_j) \| \rho_B) = F(\rho_B) \le F(\Omega(\Phi, \mathcal{A})) = C_\chi(\Phi; \mathcal{A}).$$

Then by Proposition 12.2.19, $\rho_B = \Omega(\Phi, \mathcal{A})$. If $C_\chi(\Phi; \mathcal{A}) = +\infty$, then the right-hand side of the expression in Proposition 12.2.23 is also equal to $+\infty$. Indeed, if ρ_B is a state in $\mathcal{S}(\mathbb{H}_B)$ such that

$$\sup_{\sum_j q_j \sigma_j \in \mathcal{A}} \sum_j q_j H(\Phi(\sigma_j) \| \rho_B) < +\infty,$$

then equality (12.29) is valid for arbitrarily approximating the sequence of ensembles $(\{(p_i^k, \rho_i^k)\}_{i=1}^{n(k)})_{k=1}^{+\infty}$ for the $\mathcal{A}$-constrained channel Φ that implies $C_\chi(\Phi; \mathcal{A}) < +\infty$. This proves the proposition. □

Note that the expression for $C_\chi(\Phi; \mathcal{A})$ in the above proposition can be considered as a generalization of the "mini-max formula" for χ^* in Schumacher and Westmoreland [139] to the case of an infinite-dimensional constrained channel.

Remark 12.2. Propositions 12.2.19 and 12.2.23 and Corollaries 12.2.20 and 12.2.22 do not hold without assumption of convexity of the set A. To show this it is sufficient to consider the noiseless channel $\Phi = \mathfrak{I}$ and the compact set $\mathcal{A}$, consisting of two states ρ_1 and ρ_2 such that $H(\rho_1) = H(\rho_2) < +\infty$ and $H(\rho_1 \| \rho_2) = +\infty$. In this case, $C_\chi(\Phi; \mathcal{A}) = H(\rho_1) = H(\rho_2)$, the states ρ_1 and ρ_2 are input optimal average states in the sense of Definition 12.2.10 with the different images $\Phi(\rho_1) = \rho_1$ and $\Phi(\rho_2) = \rho_2$.

12.3 Continuity of Holevo χ-capacity

In this section, the question of continuity of the χ-capacity $C_\chi(\cdot; \mathcal{A}) : \mathfrak{EQC}(A, B) \to [0, +\infty]$ as a function of extended quantum channel is considered, where $\mathcal{A}$ is a given closed subset of $\mathcal{S}(\mathbb{H}_A)$. Dealing with this question, we must choose a topology on the set $\mathfrak{QC}(A, B)$ of all quantum channels from $\mathcal{S}(\mathbb{H}_A)$ into $\mathcal{S}(\mathbb{H}_B)$. This choice is essential only in the infinite-dimensional case because all locally convex Hausdorff topologies on a finite-dimensional space are equivalent.

Let $\mathfrak{B}(\mathfrak{T}(\mathbb{H}_A), \mathfrak{T}(\mathbb{H}_B))$ be the Banach space of all bounded linear mappings from $\mathfrak{T}(\mathbb{H}_A)$ into $\mathfrak{T}(\mathbb{H}_B)$, where $\mathfrak{T}(\mathbb{H}_A)$ (resp., $\mathfrak{T}(\mathbb{H}_B)$) is the space of trace-class operators on $\mathbb{H}_A$ (resp., $\mathbb{H}_B$) under the $\|\cdot\|_1$-norm.

Definition 12.3.1. The topology on the vector space $\mathfrak{B}(\mathfrak{T}(\mathbb{H}_A), \mathfrak{T}(\mathbb{H}_B))$ defined by the family of seminorms $\{\|\Phi\|_\rho\}_{\rho\in\mathfrak{T}(\mathbb{H}_A)}$, where $\|\Phi\|_\rho := \|\Phi(\rho)\|_1$, is called the topology of strong convergence.

Since an arbitrary operator in $\mathfrak{T}(\mathbb{H}_A)$ can be represented as a linear combination of operators in $\mathcal{S}(\mathbb{H}_A)$, it is possible to consider only seminorms $\|\Phi\|_\rho := \|\Phi(\rho)\|_1$ corresponding to $\rho \in \mathcal{S}(\mathbb{H}_A)$ in the above definition. Note that a sequence $(\Phi_n)_{n=1}^{+\infty}$ of channels in $\mathfrak{QC}(A, B)$ strongly converges to a channel $\Phi \in \mathfrak{QC}(A, B)$ if and only if $\lim_{n\to+\infty} \Phi_n(\rho) = \Phi(\rho)$ for all $\rho \in \mathcal{S}(\mathbb{H}_A)$.

Theorem 12.3.2. *Let $\mathcal{A}$ be an arbitrary compact and convex subset of $\mathcal{S}(\mathbb{H}_A)$. Then $C_\chi(\cdot; \mathcal{A}) : \mathfrak{QC}(A, B) \to [0, +\infty]$ is lower semicontinuous under the strong convergence topology. Moreover, if the spaces $\mathbb{H}_A$ and $\mathbb{H}_B$ are finite- dimensional, then the Holevo χ-capacity $C_\chi(\cdot; \mathcal{A})$ is a continuous function on the set $\mathfrak{QC}(A, B)$. If $(\Phi_n)_{n=1}^{+\infty}$ is an arbitrary sequence of channels in $\mathfrak{QC}(A, B)$, converging to some channel Φ in $\mathfrak{QC}(A, B)$, then the following limit exists:*

$$\lim_{n\to+\infty} C_\chi(\Phi_n, \mathcal{A}) = C_\chi(\Phi, \mathcal{A}). \tag{12.37}$$

Proof. We first show lower semicontinuity of the χ-capacity function $C_\chi(\cdot; \mathcal{A})$ on the set $\mathfrak{QC}(A, B)$ under the strong convergence topology. Let $\epsilon > 0$ and $\{\Phi_\lambda\}$ be an arbitrary net of channels, strongly converging to the channel Φ, and let $\{(p_i, \rho_i)\}$ be an ensemble with the average $\bar{\rho}$ such that $\chi_\Phi(\{(p_i, \rho_i)\}) > C_\chi(\Phi; \mathcal{A}) - \epsilon$. By lower semicontinuity of the relative entropy $H(\cdot\|\cdot)$,

$$\begin{aligned}\liminf_\lambda \sum_i p_i H(\Phi_\lambda(\rho_i)\|\Phi_\lambda(\bar{\rho})) &\geq \sum_i p_i H(\Phi(\rho_i)\|\Phi(\bar{\rho})) \\ &= \chi_\Phi(\{p_i, \rho_i\}) > C_\chi(\Phi, \mathcal{A}) - \epsilon.\end{aligned}$$

This implies

$$\liminf_{\lambda\to+\infty} C_\chi(\Phi_\lambda; \mathcal{A}) \geq C_\chi(\Phi; \mathcal{A}).$$

It follows that

$$\liminf_{n\to+\infty} C_\chi(\Phi_n; \mathcal{A}) \geq C_\chi(\Phi, \mathcal{A}), \tag{12.38}$$

for arbitrary sequence $(\Phi_n)_{n=1}^{+\infty}$ of channels strongly converging to a channel Φ. This proves that the function $C_\chi(\cdot;\mathcal{A})$ is lower semicontinuous on the set $\mathfrak{QC}(A,B)$ under the strong convergence topology.

Now to prove the continuity of $C_\chi(\cdot;\mathcal{A})$ in the finite-dimensional case, it is sufficient to show that for the above sequence of channels

$$\limsup_{n\to+\infty} C_\chi(\Phi_n;\mathcal{A}) \le C_\chi(\Phi;\mathcal{A}). \tag{12.39}$$

For an arbitrary $\mathcal{A}$-constrained channel from $\mathfrak{QC}(A,B)$, there exists optimal ensemble consisting of $m = (\dim(\mathbb{H}))^2$ states (probably some states with zero weights). Let $\mathfrak{P}$ be the compact space of all probability distributions with m outcomes. Consider the following compact space with product topology:

$$\mathfrak{PC}^m = \mathfrak{P} \times \underbrace{\mathcal{S}(\mathbb{H}_A) \times \cdots \times \mathcal{S}(\mathbb{H})}_{m},$$

consisting of sequences $(\{p_i\}_{i=1}^m, \rho_1, \dots, \rho_m)$, corresponding to arbitrary input ensemble $\{(p_i, \rho_i)\}_{i=1}^m$ of m states. Suppose (12.39) were not true for contradiction purpose. Then without loss of generality, we may assume that

$$\lim_{n\to+\infty} C_\chi(\Phi_n, \mathcal{A}) > C_\chi(\Phi, \mathcal{A}). \tag{12.40}$$

For each n, let $\{(p_i^n, \rho_i^n)\}_{i=1}^m$ be an optimal ensemble for the $\mathcal{A}$-constrained channel Φ_n. By compactness of $\mathfrak{PC}^m$, we can choose a subsequence

$(\{p_i^{n_k}\}_{i=1}^m, \rho_1^{n_k}, \dots, \rho_m^{n_k})_{k=1}^{+\infty}$ of $(\{p_i^n\}_{i=1}^m, \rho_1^n, \dots, \rho_m^n)_{n=1}^{+\infty}$ converging to some element $(\{p_i^*\}_{i=1}^m, \rho_1^*, \dots, \rho_m^*)$ in the space $\mathfrak{PC}^m$. By definition of the product topology on $\mathfrak{PC}^m$, it means that

$$\lim_{k\to+\infty} p_i^{n_k} = p_i^* \quad \text{and} \quad \lim_{k\to+\infty} \rho_i^{n_k} = \rho_i^*, \quad i = 1, 2, \dots, m.$$

The average state of the ensemble $\{(p_i^*, \rho_i^*)\}_{i=1}^m$ is a limit of the sequence of average states of the ensembles $(\{(p_i^{n_k}, \rho_i^{n_k})\}_{i=1}^m)_{k=1}^{+\infty}$, and hence, lies in $\mathcal{A}$ (which is closed by the assumption). By continuity of the quantum entropy in the finite-dimensional case, we have

$$\lim_{k\to+\infty} C_\chi(\Phi_{n_k};\mathcal{A}) = \lim_{k\to+\infty} \chi_{\Phi_{n_k}}(\{(p_i^{n_k}, \rho_i^{n_k})\}) = \chi_\Phi(\{(p_i^*, \rho_i^*)\}) \le C_\chi(\Phi;\mathcal{A}),$$

which contradicts to (12.40). Comparing (12.38) and (12.39), we see that

$$\lim_{n\to+\infty} C_\chi(\Phi_n;\mathcal{A}) = C_\chi(\Phi;\mathcal{A}).$$

This shows that $C_\chi(\cdot;\mathcal{A})$ is a continuous on $\mathfrak{QC}(A,B)$, if $\mathbb{H}_A$ and $\mathbb{H}_B$ are both finite-dimensional.

If $(\Phi_n)_{n=1}^{+\infty}$ is an arbitrary sequence of channels in $\mathfrak{QC}(A,B)$, converging to some channel Φ in $\mathfrak{QC}(A,B)$, we want to prove that the following limit exists:

$$\lim_{n\to+\infty} \Omega(\Phi_n, \mathcal{A}) = \Omega(\Phi, \mathcal{A}),$$

where $\Omega(\Phi, \mathcal{A})$ is as defined in Proposition 12.2.19. To prove this, we note that the ensemble $\{(p_i^*, \rho_i^*)\}_{i=1}^m$ constructed above is optimal for the $\mathcal{A}$-constrained channel Φ. Hence, there exists the input optimal average state $\bar{\rho}^*$ for the $\mathcal{A}$-constrained channel Φ, which is a partial limit of the sequence $(\bar{\rho}^n)_{n=1}^{+\infty}$ of the input optimal average states for the $\mathcal{A}$-constrained channels Φ_n. Suppose (12.37) is not true. Without loss of generality, we may (by the compactness argument) assume that there exists $\lim_{n\to+\infty} \Omega(\Phi_n, \mathcal{A}) \neq \Omega(\Phi, \mathcal{A})$. By Proposition 12.2.19, this contradicts to the previous observation. This proves the theorem. □

The assumption of finite dimensionality of the Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$ is essential for the continuity of the χ-capacity $C_\chi(\cdot;\mathcal{A})$ on $\mathfrak{QC}(A,B)$ (see the second part of Theorem 12.3.2). The following example shows that generally the χ-capacity is not a continuous function of a channel, even in the stronger trace-norm topology $\|\cdot\|_1$ on the space of all channels. Although the example is a purely classical channel, it has a standard extension to a quantum one.

The following example can be found in Holevo and Shirokov [80].

Example 12.4. Consider Abelian von Neumann algebra l_∞ and its predual l_1, where l_∞ is the space of all bounded real-valued sequence $x = (x_i)_{i=1}^{+\infty}$ equipped with the sup-norm $\|x\|_{l_\infty} = \sup_i |x_i|$, and l_1 is the space of all summable real-valued sequences $x = (x_i)_{i=1}^{+\infty}$ equipped with the l_1-norm $\|x\|_{l_1} = \sum_{i=1}^{+\infty} |x_i|$.

Let $\{\Phi_n^q; n = 1,2,\dots; q \in (0,1)\}$ be the family of classical unconstrained channels from l_1 to l_1 defined by the formula

$$\Phi_n^q((x_1, x_2, \dots, x_n, \dots)) = \left((1-q)\sum_{i=1}^{+\infty} x_i, q\sum_{i=n+1}^{+\infty} x_i, qx_1, \dots, qx_n, 0, 0, \dots \right)$$

for $(x_1, x_2, \dots, x_n, \dots) \in l_1$. Defining

$$\Phi^0((x_1, x_2, \dots, x_n, \dots)) = \left(\sum_{i=1}^{+\infty} x_i, 0, 0, \dots \right),$$

we have

$$\begin{aligned}\|(\Phi_n^q - \Phi^0)((x_i)_{i=1}^{+\infty}\|_{l_1} &= q\left\| \left(-\sum_{i=1}^{+\infty} x_i, \sum_{i=n+1}^{+\infty} x_i, x_1, x_2, \dots, x_n, 0, 0, \dots \right) \right\|_{l_1} \\ &= q\left(\left|\sum_{i=1}^{+\infty} x_i\right| + \left|\sum_{i=n+1}^{+\infty} x_i\right| + |x_1| + |x_2| + \dots + |x_n| \right) \le 3q\|(x_i)_{i=1}^{+\infty}\|_{l_1}.\end{aligned}$$

Hence, $\|\Phi_n^q - \Phi^0\|_1 \to 0$ as $q \to 0$ uniformly in n.

To evaluate the χ-capacity of the channel Φ_n^q, it is sufficient to note that

$$H(\Phi_n^q(\text{any pure state})) = h_2(q) = -q\log q - (1-q)\log(1-q)$$

and

$$\begin{aligned} &H(\Phi_n^q(\text{any state})) \\ &\leq H\Big(\Phi_n^q\Big(\Big(\underbrace{\frac{1}{n+1}, \frac{1}{n+1}, \dots, \frac{1}{n+1}}_{n+1}, 0, 0, \dots\Big)\Big)\Big) \\ &= q\log(n+1) + h_2(q). \end{aligned}$$

It follows by definition that

$$\bar{C}(\Phi_n^q) = q\log(n+1), \quad q \in (0,1), n \in \mathbb{N}.$$

Take arbitrary C such that $0 < C \leq +\infty$ and choose a sequence $q(n)$ such that $\lim_{n\to+\infty} q(n) = 0$ while $\lim_{n\to+\infty} q(n)\log(n+1) = C$. Then we have $\lim_{n\to+\infty} \|\Phi_n^{q(n)} - \Phi^0\| = 0$ but $\lim_{n\to+\infty} C_\chi(\Phi_n^{q(n)}) = C > 0 = C_\chi(\Phi^0)$. Therefore, the χ-capacity $C_\chi(\cdot)$ is not continuous.

The above example demonstrates harsh discontinuity of the χ-capacity in the infinite-dimensional case. One can see that a similar discontinuity underlies Shor's construction [160] allowing to prove equivalence of different additivity properties by using channel extension and a limiting procedure.

12.4 Holevo χ-quantity and coherent information

Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel from A to B, and let $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ be its complementary channel from A to environment system E (see Definition 5.7.1). In the finite-dimensional case, *coherent information* $I_c(\rho, \Phi)$ of the channel Φ at the state ρ is defined as

$$I_c(\Phi, \rho) = H(\Phi(\rho)) - H(\hat{\Phi}(\rho)), \tag{12.41}$$

where $H(\cdot) : \mathcal{S}(\mathbb{H}) \to [0, +\infty]$ is the von Neumann entropy. However, in the infinite-dimensional case the values $H(\Phi(\rho))$ and $H(\hat{\Phi}(\rho))$ may be both infinite even for a state ρ with finite entropy. Therefore, coherent information $I_c(\rho, \Phi)$ can be defined via quantum mutual information as a function with values in $]-\infty, +\infty]$ as follows (see Proposition 12.4.1 below):

$$I_c(\rho, \Phi) = I_m(\rho, \Phi) - H(\rho).$$

Let ρ be a state in $\mathcal{S}(\mathbb{H}_A)$ with finite entropy. Then, by monotonicity of the Holevo χ-quantity, the values $\chi_\Phi(\mu)$ and $\chi_{\hat{\Phi}}(\mu)$ do not exceed $H(\rho) = \chi(\mu)$ for any measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}))$ supported by pure states with barycenter ρ.

The following proposition underlies a fundamental connection between quantum capacity and private transmission of classical information through a quantum channel. A measure for the latter is given by the difference $\chi_\Phi(\mu) - \chi_{\hat{\Phi}}(\mu)$ between the Holevo χ-quantities of the receiver and environment (eavesdropper).

Proposition 12.4.1. *Let μ be a probability measure in $\mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ supported by pure states with barycenter ρ. Then*

$$\chi_\Phi(\mu) - \chi_{\hat{\Phi}}(\mu) = I_m(\rho, \Phi) - H(\rho) = I_c(\rho, \Phi). \tag{12.42}$$

Proof. (1) We first assume that $H(\Phi(\rho)) < +\infty$, then $H(\hat{\Phi}(\rho)) < +\infty$. In this case,

$$\begin{aligned}\chi_\Phi(\mu) - \chi_{\hat{\Phi}}(\mu) &= H(\Phi(\bar{\rho}(\mu))) - \int_{\mathcal{S}(\mathbb{H})} H(\Phi(\rho))\mu(d\rho) \\ &\quad - H(\hat{\Phi}(\rho)) + \int_{\mathcal{S}(\mathbb{H})} H(\hat{\Phi}(\rho))\mu(d\rho) \\ &= I_m(\rho, \Phi) - H(\rho) = I_c(\rho, \Phi).\end{aligned}$$

(2) We now consider the general case. For a given measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$, the function $\Phi \mapsto \chi_\Phi(\mu)$ is lower semicontinuous on the set of all quantum operations equipped with the strong convergence topology for which $\Phi_n \to \Phi$ means $\Phi_n(\rho) \to \Phi(\rho)$ in $\|\cdot\|_1$-norm for all ρ. Since for an arbitrary sequence $(\Phi_n)_{n=1}^{+\infty}$ of quantum operations strongly converging to a quantum operation Φ the sequence $(\mu \circ \Phi_n^{-1})_{n=1}^{+\infty}$ weakly converges to the measure $\mu \circ \Phi^{-1}$ (this can be verified directly by using the definition of weak convergence and noting that strong convergence for sequences of quantum operations is equivalent to uniform convergence on compact subsets of $\mathcal{S}(\mathbb{H}_A)$), this follows from the lower semicontinuity of the functional $\mu \mapsto \chi(\mu)$ on the set $\mathcal{P}(\mathfrak{T}_{\leq 1}(\mathbb{H}_B))$, where

$$\mathfrak{T}_{\leq 1}(\mathbb{H}_B) = \{\mathbf{A} \in \mathfrak{T}(\mathbb{H}_B) \mid \mathbf{A} \geq 0, \mathrm{tr}[\mathbf{A}] \leq 1\}.$$

Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be an increasing sequence of finite-rank projectors in $\mathfrak{B}(\mathbb{H}_B)$ strongly converging to the identity operator $\mathbf{I}_B$. Consider the sequence of quantum operations $(\Phi_n)_{n=1}^{+\infty}$, where $\Phi_n = \Pi_n \circ \Phi$ and $\Pi_n(\sigma) = \mathbf{P}_n \sigma \mathbf{P}_n$. Then

$$\hat{\Phi}_n(\rho) = \mathrm{tr}_B[(\mathbf{P}_n \otimes \mathbf{I}_B)(\mathbf{V}\rho\mathbf{V}^*)], \quad \forall \rho \in \mathcal{S}(\mathbb{H}), \tag{12.43}$$

where $\mathbf{V} : \mathbb{H}_A \to \mathbb{H}_B \otimes \mathbb{H}_E$ is the isometry from the Stinespring representation $\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}\rho\mathbf{V}^*]$ of the channel Φ. The sequences $(\Phi_n)_{n=1}^{+\infty}$ and $(\hat{\Phi}_n)_{n=1}^{+\infty}$ converge strongly to Φ and $\hat{\Phi}$, respectively. Let $\rho = \sum_k \lambda_k |k\rangle_A\langle k|$ and $\varphi_\rho = \sum_k \sqrt{\lambda_k}|k\rangle \otimes |k\rangle$. Since $H(\rho) <$

$+\infty$ and $S(\Phi(\rho)) < +\infty$, the triangle inequality implies $S(\hat{\Phi}_n(\rho)) < +\infty$, where $S(z) = -\operatorname{tr}[z\log z]$ for $z > 0$ and $S(z) = 0$ for $z = 0$. So, we have

$$\begin{aligned} I_m(\rho,\Phi_n) &= H\big((\Phi_n \otimes \mathfrak{I}_R)(|\Phi_\rho\rangle_{AR}\langle\Phi_\rho|)\|\Phi_n(\rho)\otimes\rho\big) \\ &= -S(\hat{\Phi}_n(\rho)) + S(\Phi_n(\rho)) + a_n \\ &= -\chi_{\hat{\Phi}_n}(\mu) + \chi_{\Phi_n}(\mu) + a_n, \end{aligned} \tag{12.44}$$

where

$$a_n = -\sum_k \operatorname{tr}[\Phi_n(|k\rangle_A\langle k|)]\lambda_k \log\lambda_k, \tag{12.45}$$

and the last equality is obtained by using (12.43) and coincidence of the function $\rho \mapsto S(\Phi(\rho))$ and $\rho \mapsto S(\hat{\Phi}(\rho))$ on the set of pure states. Since the function $\Phi \mapsto I_m(\rho,\Phi)$ is lower semicontinuous (by the lower semicontinuity of relative entropy $H(\cdot\|\cdot)$ and $I_m(\rho,\Phi_n) \le I_m(\rho,\Phi)$) for all n by monotonicity of the relative entropy under the action of the quantum operation $\Pi_n \otimes \mathfrak{I}_R$, we have

$$\lim_{n\to+\infty} I_m(\rho,\Phi_n) = I_m(\rho,\Phi). \tag{12.46}$$

We will also prove that

$$\lim_{n\to+\infty}\chi_{\Phi_n}(\mu) = \chi_\Phi(\mu) \quad \text{and} \quad \lim_{n\to+\infty}\chi_{\hat{\Phi}_n}(\mu) = \chi_{\hat{\Phi}}(\mu). \tag{12.47}$$

The first relation in (12.47) follows from the lower semicontinuity of the function $\Phi \mapsto \chi_\Phi(\mu)$ established earlier and from the inequality $\chi_{\Phi_n}(\mu) \le \chi_\Phi(\mu)$ valid for all n by monotonicity of the Holevo χ-quantity under the action of the quantum operation Π_n.

To prove the second relation in (12.47), we note that (12.43) implies that $\hat{\Phi}_n(\rho) \le \hat{\Phi}(\rho)$ for all n and all $\rho \in \mathcal{S}(\mathbb{H}_A)$. Hence, Lemma 2 in Holevo and Shirokov [80] shows that

$$\chi_{\hat{\Phi}_n}(\mu) \le \chi_{\hat{\Phi}}(\mu) + f(\operatorname{tr}[\hat{\Phi}_n(\rho)]), \tag{12.48}$$

where $f(x) = -2x\log x - (1-x)\log(1-x)$, for any measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ with finite support and with barycenter ρ. Let μ be an arbitrary measure in $\mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ with barycenter ρ, and let $(\mu_k)_{k=1}^{+\infty}$ be the sequence of measures with finite support with the same barycenter constructed in the proof of Lemma 1 in Holevo and Shrikov [80], which weakly converges to the measure μ. Validity of inequality (12.48) for the measure μ is derived from its validity for all measures μ_k by using lower semicontinuity of the function $\mu \mapsto \chi_{\Phi_n}(\mu)$ and the inequality $\chi_\Phi(\mu_k) \le \chi_\Phi(\mu)$, which is valid for all k by the construction of the sequence $(\mu_k)_{k=1}^{+\infty}$ and convexity of the relative entropy.

Inequality (12.48) and lower semicontinuity of the function $\Phi \mapsto \chi_\Phi(\mu)$ imply the second relation in (12.47).

Since the sequence $(a_n)_{n=1}^{+\infty}$ defined in (12.45) obviously tends to $H(\rho)$, relations (12.44), (12.46) and (12.47) imply (12.42). This proves the proposition. □

12.5 Additivity of Holevo χ-capacities

Let $\mathbb{H}_A$, $\mathbb{H}_B$, $\mathbb{K}_A$ and $\mathbb{K}_B$ be separable complex Hilbert spaces, and let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and $\Psi : \mathcal{S}(\mathbb{K}_A) \to \mathcal{S}(\mathbb{K}_B)$ be two quantum channels from $\mathbb{H}_A$ to $\mathbb{H}_B$ and from $\mathbb{K}_A$ to $\mathbb{K}_B$, respectively.

This section is concerned with the additivity question of the Holevo χ-capacity $C_\chi(\cdot)$ as a function of quantum channels. That is, whether or not $C_\chi(\Phi \otimes \Psi) = C_\chi(\Phi) + C_\chi(\Psi)$. Roughly speaking, the Holevo χ-capacity $C_\chi(\cdot)$ defined in (12.10) is said to be additive if

$$C_\chi(\Phi \otimes \Psi) = C_\chi(\Phi) + C_\chi(\Psi), \quad \forall \Phi \in \mathfrak{QC}(\mathbb{H}_A, \mathbb{H}_B) \text{ and } \forall \Psi \in \mathfrak{QC}(\mathbb{K}_A, \mathbb{K}_B).$$

Although we use the same notation $C_\chi(\cdot)$ for all Holevo χ-capacities here, the left-hand side $C_\chi(\Phi \otimes \Psi)$ is the χ-capacity for the tenor product channel $\Phi \otimes \Psi$ defines as

$$C_\chi(\Phi \otimes \Psi) = \sup_{\nu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_{AB}))} \chi_{\Phi\otimes\Psi}(\nu), \tag{12.49}$$

whereas $C_\chi(\Phi)$ and $C_\chi(\Psi)$ on the right-hand side are defined in (12.10), etc.

Other related concepts such as subadditivity and superadditivity of Holevo χ-capacity are to be defined in Definition 12.5.3.

Proving the additivity of the Holevo χ-capacity was one of the most important problems in quantum information theory over the past decade, because additivity implies that entangled inputs between the two channels Φ and Ψ cannot enhance the total rate of information transmitted through Φ and Ψ individually. This is because if the Holevo χ-capacity is additive, then infinitely repeated use of a memoryless quantum channel Φ yields its asymptotic average classical capacity $C(\Phi)$ of the quantum channel equals (see Proposition 13.3.3 for the relation between classical capacity $C(\cdot)$ and the Holevo χ-capacity $C_\chi(\cdot)$ for memoryless channels):

$$C(\Phi) = \lim_{n\to+\infty} \frac{1}{n} C_\chi(\Phi^{\otimes n}) = \lim_{n\to+\infty} \frac{1}{n} \sum_{i=1}^{n} C_\chi(\Phi_i) = C_\chi(\Phi),$$

where $\Phi = \Phi_i$ for all $= 1, 2, \ldots$. In this case, the quantum capacity $C(\Phi)$ is exactly $C_\chi(\Phi)$ and no capacity is gained from entanglement of input and output systems (see Section 13.4 for entanglement assisted channels that improves the information transmission rate).

If the channel Φ has additive Holevo χ-capacity, then computation of its classical capacity $C(\Phi)$ will be significantly simplified and less taxing by only computing

$C_\chi(\Phi)$. For this reason alone, one might wish that the Holevo χ-capacity is always additive for all channels. Unfortunately, while additivity of the Holevo χ-capacity $C_\chi(\cdot)$ has been shown to be true in depolarizing, unital and entanglement breaking channels (see, e. g., King [97]), the superadditivity of the Holevo χ-capacity in certain higher-dimensional quantum channels was first given by Hastings [60] who showed the minimum output entropy of the channel Φ and its conjugate channel $\bar{\Phi}$ satisfies the following strict inequality:

$$\min_{\rho,\sigma} H\big((\Phi \otimes \bar{\Phi})(\rho \otimes \sigma)\big) < \min_{\rho} H\big(\Phi(\rho)\big) + \min_{\sigma} H\big(\bar{\Phi}(\sigma)\big),$$

where the minimum is over all rank one pure states ρ and σ. This disproves the additivity conjecture that Holevo χ-capacity is always additive. One of consequences of the additivity violation of minimum output entropy is that Holevo capacity is not in general additive either.

Hastings' counterexample to the additivity conjecture is given in the following subsection without a proof.

12.5.1 Hastings' counterexamples to additivity

The first counterexample to the additivity conjecture was given by Hastings [60]. Other counterexamples are also given by King [97] (see also the references contained therein).

Let N and M be finite positive integers. For $i = 1, \ldots, N$, pick $l_i \geq 0$ independently from a probability distribution

$$P(l_i) \propto l_i^{2M-1} \exp(-MNl_i^2), \quad i = 1, 2, \ldots, N,$$

where the proportionality constant is chosen such that $\int_0^{+\infty} P(l)dl = 1$. This distribution is the same as that of the length of a random vector chosen from a Gaussian distribution in M complex dimensions. Then define

$$L = \sqrt{\sum_{i=1}^{N} l_i^2}.$$

Then we set

$$P_i = l_i^2/L^2.$$

Define a pair of channels Φ and $\bar{\Phi}$, which are complex conjugates of each other. Each channel acts by randomly choosing a unitary from a small set of unitaries $\mathbf{U}_i$, $i = 1, \ldots, N$, and applying that to ρ. This models a situation in which the unitary evolution of the system is determined by an unknown state of the environment. We define

$$\Phi(\rho) = \sum_{i=1}^{N} P_i \mathbf{U}_i^* \rho \mathbf{U}_i = \sum_{i=1}^{N} \frac{l_i^2}{L^2} \mathbf{U}_i^* \rho \mathbf{U}_i, \tag{12.50}$$

and

$$\bar{\Phi}(\rho) = \sum_{i=1}^{N} P_i \bar{\mathbf{U}}_i^* \rho \bar{\mathbf{U}}_i = \sum_{i=1}^{N} \frac{l_i^2}{L^2} \bar{\mathbf{U}}_i^* \rho \bar{\mathbf{U}}_i. \tag{12.51}$$

The proof of the Hastings counterexample below can be found in [60] and is omitted here.

Theorem 12.5.1 (Hastings' counterexample [60]). *For sufficiently large N and for sufficiently large M that depends on N, there is a nonzero probability that a random choice of $\mathbf{U}_i$ from the Haar measure and of P_i that define channels Φ and $\bar{\Phi}$ in* (12.50) *and* (12.51), *respectively. Then*

$$\begin{aligned} \min_{\rho,\sigma} H\big((\Phi \otimes \bar{\Phi})(\rho \otimes \sigma)\big) &< \min_{\rho} H\big(\Phi(\rho)\big) + \min_{\sigma} H\big(\bar{\Phi}(\sigma)\big) \\ &= 2 \min_{\rho} H\big(\Phi(\rho)\big), \end{aligned} \tag{12.52}$$

where the minimum is over all pure states of rank one.

12.5.2 Additivity of unconstrained χ-quantity

Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel from system A to system B. This subsection explores various conditions under which the unconstrained Holevo χ-capacity of channel Φ is additive or nonadditive.

Since unconstrained Holevo χ-capacity $C_\chi(\Phi)$ of channel Φ is defined in terms of its Holevo χ- quantity $\chi_\Phi(\cdot)$ (see (12.10)) as

$$C_\chi(\Phi) = \sup_{\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))} \chi_\Phi(\mu),$$

to address the additivity question of $C_\chi(\Phi)$, we first need to look at the additivity of its corresponding Holevo χ quantity $\chi_\Phi(\cdot)$, which depends on the dimensionality of the channel Φ.

We classify the type of the channel Φ based on the dimensionality of its input system A and output system B below.

Definition 12.5.2. The channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is said to be an:
1. *FF* channel if the input space $\mathbb{H}_A$ and the output space $\mathbb{H}_B$ are both finite-dimensional.
2. *FI* channel if the input space $\mathbb{H}_A$ is finite-dimensional and the output space $\mathbb{H}_B$ is infinite-dimensional.

3. *IF* channel if the input space $\mathbb{H}_A$ is infinite-dimensional and the output space $\mathbb{H}_B$ is finite-dimensional.
4. *II* channel if its input space $\mathbb{H}_A$ and its output space $\mathbb{H}_B$ are both infinite-dimensional.

A similar definition also applies to the channel $\Psi : \mathcal{S}(\mathbb{K}_A) \to \mathcal{S}(\mathbb{K}_B)$, where $\mathbb{K}_A$ and $\mathbb{K}_B$ are the Hilbert spaces representing the input and output systems, respectively, of the channel Ψ.

Recall that the Holevo χ-quantity for a channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is the function $\chi_\Phi(\cdot) : \mathcal{P}(\mathcal{S}(\mathbb{H}_A)) \to [0, +\infty]$ (see Definition 12.1.2) defined by

$$\chi_\Phi(\mu) = \int\limits_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho)\|\Phi(\bar{\rho}))\mu(d\rho), \quad \forall \mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A)),$$

where $\mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ is the space of Borel probability measures on $\mathcal{S}(\mathbb{H}_A)$, and $\bar{\rho}(\mu) := \int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu(d\rho)$ is the barycenter (or the average state) of the measure $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$.

Definition 12.5.3. The Holevo χ-quantity $\chi_{\Phi\otimes\Psi}(\mu \otimes \nu)$ is said to be additive if, for all $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$, $\nu \in \mathcal{P}(\mathcal{S}(\mathbb{K}_A))$,

$$\chi_{\Phi\otimes\Psi}(\mu \otimes \nu) = \chi_\Phi(\mu) + \chi_\Psi(\nu). \tag{12.53}$$

The Holevo χ-quantity $\chi_{\Phi\otimes\Psi}(\mu \otimes \nu)$ is said to be subadditive if, for all $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$, $\nu \in \mathcal{P}(\mathcal{S}(\mathbb{K}_A))$,

$$\chi_{\Phi\otimes\Psi}(\mu \otimes \nu) \le \chi_\Phi(\mu) + \chi_\Psi(\nu). \tag{12.54}$$

The Holevo χ-quantity $\chi_{\Phi\otimes\Psi}(\mu \otimes \nu)$ is said to be superadditivity if, for all $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$, $\nu \in \mathcal{P}(\mathcal{S}(\mathbb{K}_A))$,

$$\chi_{\Phi\otimes\Psi}(\mu \otimes \nu) \ge \chi_\Phi(\mu) + \chi_\Psi(\nu). \tag{12.55}$$

We shall use the following chain properties of the Holevo χ-quantity.

Proposition 12.5.4. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and $\Psi : \mathcal{S}(\mathbb{H}_B) \to \mathcal{S}(\mathbb{H}_C)$ be channels from $\mathbb{H}_A$ to $\mathbb{H}_B$ and from $\mathbb{H}_B$ to $\mathbb{H}_C$, respectively. Then*

$$\chi_{\Psi\circ\Phi}(\rho) \le \chi_\Phi(\rho) \quad \textit{and} \quad \chi_{\Psi\circ\Phi}(\rho) \le \chi_\Psi(\Phi(\rho)), \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A).$$

Proof. The first inequality follows from the monotonicity property of the relative entropy $H(\cdot\|\cdot)$ (see Lemma 9.1.3) and the definition of the Holevo χ-quantity (see (9.2)) and the fact that $\chi_\Phi(\rho) = C_\chi(\Phi; \{\rho\})$, while the second one is a direct corollary of the Definition 12.1.2 of the Holevo χ-quantity $\chi_{\Psi\circ\Phi}(\{\rho\})$. This proves the proposition. □

Suppose that $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is either an arbitrary *FI* or an arbitrary *II* channel. That is, $\dim(\mathbb{H}_A) < +\infty$ and $\dim(\mathbb{H}_B) = +\infty$ or $\dim(\mathbb{H}_A) = +\infty$ and $\dim(\mathbb{H}_B) = +\infty$. In this case, let $(\mathbf{P}_B^n)_{n=1}^{+\infty}$ be a sequence of finite rank projectors on $\mathbb{H}_B$ strongly converging to $\mathbf{I}_{\mathbb{H}_B} := \mathbf{I}_B$ (the identity operator on $\mathbb{H}_B$), and let $\mathbb{H}_B^n = \mathbf{P}_B^n(\mathbb{H}_B)$. Consider the sequence of channels $(\Phi_n)_{n=1}^{+\infty}$ from $\mathcal{S}(\mathbb{H}_A)$ into $\mathcal{S}(\mathbb{H}_B^n) = \mathcal{S}(\mathbb{H}_B^{\prime n} \oplus \mathbb{H}_B^{\prime\prime n}) \subset \mathcal{S}(\mathbb{H}_B)$ defined by

$$\Phi_n(\rho) = \mathbf{P}_B^n \Phi(\rho) \mathbf{P}_B^n + \mathrm{tr}[(\mathbf{I}_B - \mathbf{P}_B^n)\Phi(\rho)]\tau_n, \tag{12.56}$$

where τ_n is a pure state in some finite-dimensional subspace $\mathbb{H}_B^{\prime\prime n}$ of $\mathbb{H}_B \ominus \mathbb{H}_B^{\prime n}$ (note: $\mathbb{H}_B \ominus \mathbb{H}_B^{\prime n}$ is the subspace of $\mathbb{H}_B$ such that $(\mathbb{H}_B \ominus \mathbb{H}_B^{\prime n}) \oplus \mathbb{H}_B^{\prime n} = \mathbb{H}_B$). If $\dim(\mathbb{H}_B) < +\infty$, we will assume that $\Phi_n = \Phi$ for all n. Note that for arbitrary *FI*-channel Φ the corresponding channel Φ_n is a *FF*-channel for all n, since $\dim(\mathbb{H}_B^n) < +\infty$.

For arbitrary channel $\Psi : \mathcal{S}(\mathbb{K}_A) \to \mathcal{S}(\mathbb{K}_B)$, we will consider the sequences Φ_n and $\Phi_n \otimes \Psi$ of channels as approximations for the channels Φ and $\Phi \otimes \Psi$, respectively. Despite the discontinuity of the Holevo χ-capacity as a function of a channel in the infinite-dimensional case (see Theorem 12.3.2), the following result (due originally to Shirokov [142]) is valid. The result states that if the subadditivity holds for an *FF* channel Φ_n tensor product with another channel Ψ, then subadditivity also holds for the *IF* Φ tensor product with Φ by using strong convergence of the projection operators $\{\mathbf{P}_B^n\}_{n=1}^{+\infty}$ to $\mathbf{I}_B$.

Lemma 12.5.5. *Let Φ and Ψ be arbitrary channels from $\mathbb{H}_A$ to $\mathbb{H}_B$ and from $\mathbb{K}_A$ to $\mathbb{K}_B$, respectively. If subadditivity of the Holevo χ-quantity holds for the channel Φ_n defined by* (12.56) *and the channel Ψ for all n, then subadditivity of the χ-function holds for the channels Φ and Ψ.*

Proof. The channel Φ_n can be represented as the composition $\Pi_n \circ \Phi$ of the channel Φ with the channel $\Pi_n : \mathcal{S}(\mathbb{H}_B) \to \mathcal{S}(\mathbb{H}_B^{\prime n} \oplus \mathbb{H}_B^{\prime\prime n})$ defined

$$\Pi_n(\rho_B) = \mathbf{P}_B^n \rho_B \mathbf{P}_B^n + (\mathrm{tr}[(\mathbf{I}_B - \mathbf{P}_B^n)\rho_B])\tau_n.$$

Proposition 12.5.4 implies that

$$\chi_{\Phi_n}(\rho) = \chi_{\Pi_n \circ \Phi}(\rho) \le \chi_\Phi(\rho), \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A) \text{ and } \forall n \in \mathbb{N}.$$

Since

$$\lim_{n\to+\infty} \Phi_n(\rho) = \Phi(\rho), \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A),$$

it follows from Theorem 12.3.2 that

$$\liminf_{n\to+\infty} \chi_{\Phi_n}(\rho) \ge \chi_\Phi(\rho), \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A).$$

The above two inequalities imply that

$$\lim_{n\to+\infty}\chi_{\Phi_n}(\rho)=\chi_{\Phi}(\rho),\quad \forall\rho\in\mathcal{S}(\mathbb{H}_A). \tag{12.57}$$

It is easy to see that

$$(\Phi_n\otimes\Psi)(\omega)=(\mathbf{P}_B^n\otimes\mathbf{I}_{\mathbb{K}_B})(\Phi\otimes\Psi)(\omega)(\mathbf{P}_B^n\otimes\mathbf{I}_{\mathbb{K}_B}) \\ +\tau_n\otimes\mathrm{tr}_B[((\mathbf{I}_B-\mathbf{P}_B^n)\otimes\mathbf{I}_{\mathbb{K}_B})((\Phi\otimes\Psi)(\omega))],\quad \forall\omega\in\mathcal{S}(\mathbb{H}_A\otimes\mathbb{K}_A).$$

Hence,

$$\lim_{n\to+\infty}(\Phi_n\otimes\Psi)(\omega)=(\Phi\otimes\Psi)(\omega),\quad \forall\omega\in\mathcal{S}(\mathbb{H}_A\otimes\mathbb{K}_A),$$

and by Theorem 12.3.2 we have

$$\liminf_{n\to+\infty}\chi_{\Phi_n\otimes\Psi}(\omega)\geq\chi_{\Phi\otimes\Psi}(\omega),\quad \forall\omega\in\mathcal{S}(\mathbb{H}_A\otimes\mathbb{K}_A). \tag{12.58}$$

By the assumption, we have for all $n\in\mathbb{N}$,

$$\chi_{\Phi_n\otimes\Psi}\leq\chi_{\Phi_n}(\omega_{\mathbb{H}_A})+\chi_{\Psi}(\omega_{\mathbb{K}_A}),\quad \forall\omega\in\mathcal{S}(\mathbb{H}_A\otimes\mathbb{K}_A).$$

This, (12.57) and (12.58) imply that

$$\chi_{\Phi\otimes\Psi}(\omega)\leq\chi_{\Phi}(\omega_{\mathbb{H}_A})+\chi_{\Psi}(\omega_{\mathbb{K}_A}),\quad \forall\omega\in\mathcal{S}(\mathbb{H}_A\otimes\mathbb{K}_A).$$

This proves the lemma. □

Proposition 12.5.6. *Subadditivity of the Holevo χ-quantity for all FF-channels implies subadditivity of the Holevo χ-quantity for all FI-channels.*

Proof. This can be proved by double application of Lemma 12.5.5. First, we prove the subadditivity of the Holevo χ-quantity for any two channels, when one of them is of *FI*-type while another is of *FF*-type. Second, we remove the *FF* restriction from the last channel. This proves the proposition. □

Now we will turn to channels with the infinite-dimensional input quantum system $\mathbb{H}_A$. We will use the following notion of the subchannel.

Definition 12.5.7. The restriction of a channel $\Phi:\mathcal{S}(\mathbb{H}_A)\to\mathcal{S}(\mathbb{H}_B)$ to the set of states with support contained in a subspace $\mathbb{H}_0$ of the space $\mathbb{H}_A$ is called subchannel Φ_0 of the channel Φ, corresponding to the subspace $\mathbb{H}_0$.

It is easy to see that subadditivity of the Holevo χ-quantity for the channels Φ and Ψ implies subadditivity of the Holevo χ-quantity for arbitrary subchannels Φ_0 and Ψ_0 of the channels Φ and Ψ. The properties of the Holevo χ-quantity established in Section 12.1 make it possible to prove the following important result.

Proposition 12.5.8. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and $\Psi : \mathcal{S}(\mathbb{K}_A) \to \mathcal{S}(\mathbb{K}_B)$ be arbitrary channels. Subadditivity of the Holevo χ-quantity for any two FI-subchannels of the channels Φ and Ψ implies subadditivity of the Holevo χ-quantity for the channels Φ and Ψ.*

Proof. It is sufficient to consider the case $\dim(\mathbb{H}_A) = +\infty$ and $\dim(\mathbb{K}_{\mathbb{A}}) \le +\infty$. Let ω be an arbitrary state in $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$. Let $\{|\phi_k\rangle\}_{k=1}^{+\infty}$ and $\{|\psi_k\rangle\}_{k=1}^{\dim(\mathbb{K}_A)}$ be orthonormal bases (ONB) of eigenvectors of the compact positive operators $\omega_{\mathbb{H}}$ and $\omega_{\mathbb{K}}$ such that the corresponding sequences of eigenvalues are nonincreasing. Let $\mathbf{P}_n = \sum_{k=1}^{n} |\phi_k\rangle_{\mathbb{H}_A}\langle\phi_k|$ and $\mathbf{Q}_n = \sum_{k=1}^{n} |\psi_k\rangle_{\mathbb{K}_A}\langle\psi_k|$. In the case $\dim(\mathbb{K}_A) < +\infty$, we will assume $\mathbf{Q}_n = \mathbf{I}_{\mathbb{K}}$ for all $n \ge \dim(\mathbb{K}_A)$. The nondecreasing sequences $(\mathbf{P}_n)_{n=1}^{+\infty}$ and $(\mathbf{Q}_n)_{n=1}^{+\infty}$ of finite-rank projectors converge to $\mathbf{I}_{\mathbb{H}_A}$ and to $\mathbf{I}_{\mathbb{K}_A}$ correspondingly in the strong operator topology. Let $\mathbb{H}_{n,A} = \mathbf{P}_n(\mathbb{H}_A)$ and $\mathbb{K}_{n,A} = \mathbf{Q}_n(\mathbb{K}_A)$. Consider the following sequence of states in $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$:

$$\omega_n = \big(\mathrm{tr}[(\mathbf{P}_n \otimes \mathbf{Q}_n)\omega]\big)^{-1}(\mathbf{P}_n \otimes \mathbf{Q}_n)\omega \cdot (\mathbf{P}_n \otimes \mathbf{Q}_n),$$

which are well-defined for all n by the choice of the projectors $\mathbf{P}_n$ and $\mathbf{Q}_n$. Since

$$\lim_{n\to+\infty} \omega_n = \omega, \tag{12.59}$$

Theorem 12.3.2 implies that

$$\liminf_{n\to+\infty} \chi_{\Phi\otimes\Psi}(\omega_n) \ge \chi_{\Phi\otimes\Psi}(\omega). \tag{12.60}$$

The next part of the proof is based on the following operator inequalities:

$$\lambda_n \omega_{n,\mathbb{H}} \le \omega_{\mathbb{H}} \quad \text{and} \quad \lambda_n \omega_{n,\mathbb{K}} \le \omega_{\mathbb{K}}, \tag{12.61}$$

where $\lambda_n = \mathrm{tr}[(\mathbf{P}_n \otimes \mathbf{Q}_n)\omega]$, $\omega_{n,\mathbb{H}} = \mathrm{tr}_{\mathbb{K}}[\omega_n]$ and $\omega_{n,\mathbb{K}} = \mathrm{tr}_{\mathbb{H}}[\omega_n]$. We prove the first inequality of (12.61) as follows. By the choice of $\mathbf{P}_n$ and due to the fact that $\mathrm{supp}(\omega_{n,\mathbb{H}}) \subseteq \mathbb{H}_n$, it is sufficient to show that $\lambda_n\omega_n \le \mathbf{P}_n\omega$. Let $\varphi \in \mathbb{H}_{n,A}$. By definition of partial trace,

$$\begin{aligned}
\langle\varphi|\lambda_n\omega_{\mathbb{H}_{n,A}}|\varphi\rangle_{\mathbb{H}_A} &= \sum_{k=1}^{\dim(\mathbb{K}_A)} \langle\varphi\otimes\psi_k|(\mathbf{P}_n \otimes \mathbf{Q}_n)\omega(\mathbf{P}_n \otimes \mathbf{Q}_n)|\varphi\otimes\psi_k\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A} \\
&= \sum_{k=1}^{m} \langle\varphi\otimes\psi_k|\omega|\varphi\otimes\psi_k\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A} \\
&\le \sum_{k=1}^{\dim(\mathbb{K}_A)} \langle\varphi\otimes\psi_k|\omega|\varphi\otimes\psi_k\rangle_{\mathbb{H}_A\otimes\mathbb{K}_A} = \langle\varphi|\omega_{\mathbb{H}_A}|\varphi\rangle_{\mathbb{H}_A},
\end{aligned}$$

where $m = \min\{n, \dim(\mathbb{K}_A)\}$. The second inequality is proved by the same way. By using (12.59) and due to (12.61), we obtain

$$\lim_{n\to+\infty} \chi_\Phi(\omega_{n,\mathbb{H}_A}) = \chi_\Phi(\omega_{\mathbb{H}_A}) \quad \text{and} \quad \lim_{n\to+\infty} \chi_\Psi(\omega_{n,\mathbb{K}_A}) = \chi_\Psi(\omega_{\mathbb{K}_A}). \tag{12.62}$$

For each n, the $\{\omega_{n,\mathbb{H}_A}\}$-constrained channel Φ and the $\{\omega_{n,\mathbb{K}_A}\}$-constrained channel Ψ can be considered as *FI*-subchannels of the channels Φ and Ψ corresponding to the subspaces $\mathbb{H}_{n,A}$ and $\mathbb{K}_{n,A}$. Hence, by the assumption,

$$\chi_{\Phi\otimes\Psi}(\omega_n) \leq \chi_\Phi(\omega_{n,\mathbb{H}_A}) + \chi_\Psi(\omega_{n,\mathbb{K}_A}), \quad \forall n \in \mathbb{N}.$$

This, (12.60) and (12.62) imply

$$\chi_{\Phi\otimes\Psi}(\omega) \leq \chi_\Phi(\omega_{\mathbb{H}_A}) + \chi_\Psi(\omega_{\mathbb{K}_A}), \quad \forall \omega \in \mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A).$$

This proves the proposition. □

12.5.3 Tensor product of constrained channels

Let $\mathbb{H}_A$, $\mathbb{H}_B$, $\mathbb{K}_A$ and $\mathbb{K}_B$ be separable complex Hilbert spaces. Let $\mathcal{H}$ and $\mathcal{K}$ be closed subsets of $\mathcal{S}(\mathbb{H}_A)$ and $\mathcal{S}(\mathbb{K}_A)$, respectively. We define $\mathcal{H}\otimes\mathcal{K}$, the tensor product of $\mathcal{H}$ and $\mathcal{K}$, as

$$\mathcal{H}\otimes\mathcal{K} = \{\omega \in \mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A) \mid \omega_{\mathbb{H}_A} := \mathrm{tr}_{\mathbb{K}_A}[\omega] \in \mathcal{H}, \omega_{\mathbb{K}_A} := \mathrm{tr}_{\mathbb{H}_A}[\omega] \in \mathcal{K}\}. \tag{12.63}$$

Let Φ be an $\mathcal{H}$-constrained channel from $\mathbb{H}_A$ to $\mathbb{H}_B$ and Ψ be a $\mathcal{K}$-constrained channel from $\mathbb{K}_A$ to $\mathbb{K}_B$. The channel $\Phi\otimes\Psi$ is said to be a $\mathcal{H}\otimes\mathcal{K}$-constrained channel from $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$ to $\mathcal{S}(\mathbb{H}_B \otimes \mathbb{K}_B)$ if the following requirements are satisfied for all $\omega \in \mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$: (i) $\bar{\omega}_{\mathbb{H}_A} := \mathrm{tr}_{\mathbb{K}_A}[\bar{\omega}] \in \mathcal{H}$ and (ii) $\bar{\omega}_{\mathbb{K}_A} := \mathrm{tr}_{\mathbb{H}_A}[\bar{\omega}] \in \mathcal{K}$, where $\bar{\omega}$ is the average state of an ensemble $\{q_i, \omega_i\}$.

We have the following result on $\mathcal{H}\otimes\mathcal{K}$ in terms of its components $\mathcal{H}$ and $\mathcal{K}$.

Lemma 12.5.9. *The following topological characterizations of $\mathcal{H}\otimes\mathcal{K}$ in terms of $\mathcal{H}$ and $\mathcal{K}$ holds:*

1. *The set $\mathcal{H}\otimes\mathcal{K}$ is a convex subset of $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$ if and only if the sets $\mathcal{H}$ and $\mathcal{K}$ are convex subsets of $\mathcal{S}(\mathbb{H}_A)$ and of $\mathcal{S}(\mathbb{K}_A)$, respectively.*
2. *The set $\mathcal{H}\otimes\mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$ if and only if the sets $\mathcal{H}$ and $\mathcal{K}$ are compact subsets of $\mathcal{S}(\mathbb{H}_A)$ and of $\mathcal{S}(\mathbb{K}_A)$, respectively.*

Proof. 1. The first statement of this lemma is trivial.

2. We want to prove second statement of the lemma that $\mathcal{H}\otimes\mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$ if and only if $\mathcal{H}$ is a compact of $\mathcal{S}(\mathbb{H}_A)$ and $\mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{K}_A)$. First, we assume that $\mathcal{H}\otimes\mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$. Note that

$$\mathcal{H} = \{\mathrm{tr}_{\mathbb{K}_A}[\omega] \mid \omega \in \mathcal{H}\otimes\mathcal{K}\} := \mathrm{tr}_{\mathbb{K}_A}[\mathcal{H}\otimes\mathcal{K}]$$

and

$$\mathcal{K} = \{\mathrm{tr}_{\mathbb{H}_A}[\omega] \mid \omega \in \mathcal{H} \otimes \mathcal{K}\} := \mathrm{tr}_{\mathbb{H}_A}[\mathcal{H} \otimes \mathcal{K}].$$

From the above relations, it is clear that the partial trace $\mathrm{tr}_{\mathbb{K}_A}[\cdots]$ and $\mathrm{tr}_{\mathbb{H}_A}[\cdots]$ are continuous functions from $\mathcal{H}\otimes\mathcal{K}$ onto $\mathcal{H}$ and $\mathcal{K}$, respectively. By using a well-known fact in topology that the image of a continuous map on a compact set is compact (see Rudin [133]), we can easily conclude that the compactness of $\mathcal{H} \otimes \mathcal{K}$ implies the compactness of $\mathcal{H}$ in $\mathcal{S}(\mathbb{H})$ and the compactness of $\mathcal{K}$ in $\mathcal{S}(\mathbb{K})$. Conversely, we assume that $\mathcal{H}$ and $\mathcal{K}$ are compact subsets of $\mathcal{S}(\mathbb{H})$ and $\mathcal{S}(\mathbb{K})$, respectively, and we want to show that $\mathcal{H} \otimes \mathcal{K}$ is a compact subset of $\mathcal{S}(\mathbb{H}\otimes\mathbb{K})$. To prove this, let $\mathcal{H}$ and $\mathcal{K}$ be compact subsets of $\mathcal{S}(\mathbb{H})$ and $\mathcal{S}(\mathbb{K})$, respectively. By Proposition 3.2.2, for arbitrary $\epsilon > 0$ there exist finite rank projectors $\mathbf{P}_\epsilon$ and $\mathbf{Q}_\epsilon$ such that

$$\mathrm{tr}[\mathbf{P}_\epsilon \rho] > 1 - \epsilon, \quad \forall \rho \in \mathcal{H} \quad \text{and} \quad \mathrm{tr}[\mathbf{Q}_\epsilon \sigma] > 1 - \epsilon, \quad \forall \sigma \in \mathcal{K}.$$

Since $\omega_{\mathbb{H}} := \mathrm{tr}_{\mathbb{K}}[\omega] \in \mathcal{H}$ and $\omega_{\mathbb{K}} := \mathrm{tr}_{\mathbb{H}}[\omega] \in \mathcal{K}$ for arbitrary $\omega \in \mathcal{H} \otimes \mathcal{K}$, we have

$$\begin{aligned}\mathrm{tr}[(\mathbf{P}_\epsilon \otimes \mathbf{Q}_\epsilon)\omega] &= \mathrm{tr}[(\mathbf{P}_\epsilon \otimes \mathbf{I}_{\mathbb{K}})\omega] - \mathrm{tr}[(\mathbf{P}_\epsilon \otimes (\mathbf{I}_{\mathbb{K}} - \mathbf{Q}_\epsilon))\omega] \\ &\geq \mathrm{tr}[\mathbf{P}_\epsilon \omega_{\mathbb{H}}] - \mathrm{tr}[(\mathbf{I}_{\mathbb{K}} - \mathbf{Q}_\epsilon)\omega_{\mathbb{K}}] > 1 - 2\epsilon.\end{aligned}$$

By Proposition 3.2.2, compactness of $\mathcal{H}$ and $\mathcal{K}$ imply compactness of the set $\mathcal{H} \otimes \mathcal{K}$. This proves the lemma. ☐

12.5.4 Additivity of constrained Holevo χ-capacity

Let $\mathcal{H}$ and $\mathcal{K}$ be closed subsets of $\mathcal{S}(\mathbb{H}_A)$ and $\mathcal{S}(\mathbb{K}_A)$, respectively. The following result, due originally to Shirokov [141], answers the question of the additivity of the Holevo χ-capacity of $\mathcal{H} \otimes \mathcal{K}$-constrained channel $\Phi \otimes \Psi$. More precisely, investigations are focused on the conditions under which the following equality holds:

$$C_\chi(\Phi \otimes \Psi; \mathcal{H} \otimes \mathcal{K}) = C_\chi(\Phi; \mathcal{H}) + C_\chi(\Psi; \mathcal{K}). \tag{12.64}$$

Note that when $\mathcal{H} = \mathcal{S}(\mathbb{H}_A)$ and $\mathcal{K} = \mathcal{S}(\mathbb{K}_A)$, the $\mathcal{H} \otimes \mathcal{K}$-constrained channel $\Phi \otimes \Psi$ becomes a unconstrained one and the results obtained in this subsection automatically reduce to that of the unconstrained case.

The following lemma is needed to answer the additivity question of the constrained χ-capacity.

Lemma 12.5.10. *Let $\Omega(\Phi; \mathcal{H})$ and $\Omega(\Psi; \mathcal{K})$ be the output optimal average states for the $\mathcal{H}$-constrained channel Φ and the $\mathcal{K}$-constrained channel Ψ, respectively. Then $\Omega(\Phi; \mathcal{H}) \otimes \Omega(\Psi; \mathcal{K})$ is the output optimal average state for the $\mathcal{H} \otimes \mathcal{K}$-constrained channel $\Phi \otimes \Psi$.*

Proof. Let $(\{p_i^k, \rho_i^k\})_{k=1}^{+\infty}$ and $(\{q_j^k, \sigma_j^k\})_{k=1}^{+\infty}$ be approximating sequences of ensembles for the $\mathcal{H}$-constrained channel Φ and $\mathcal{K}$-constrained channel Ψ, respectively. For each k, let $\bar{\rho}^k = \sum_i p_i^k \rho_i^k$ and $\bar{\sigma}^k = \sum_j q_j^k \sigma_j^k$. By Proposition 12.2.19, the sequences $(\Phi(\bar{\rho}^k))_{k=1}^{+\infty}$ and $(\Psi(\bar{\sigma}^k))_{k=1}^{+\infty}$ converge to $\Omega(\Phi;\mathcal{H})$ and to $\Omega(\Psi;\mathcal{K})$, respectively. Consequently, the sequence of ensembles $(\{p_i^k q_j^k, \rho_i^k \otimes \sigma_j^k\})_{k=1}^{+\infty}$ is an approximating optimal sequence for the $\mathcal{H}\otimes\mathcal{K}$-constrained channel $\Phi\otimes\Psi$. By Proposition 12.2.19, the limit $\Omega(\Phi,\mathcal{H})\otimes\Omega(\Psi,\mathcal{K})$ of the sequence $(\Phi(\bar{\rho}^k) \otimes \Psi(\bar{\sigma}^k))_{k=1}^{+\infty}$ is the output optimal average states of the $\mathcal{H} \otimes \mathcal{K}$-constrained channel $\Phi \otimes \Psi$. This proves the lemma. □

The results above enable us to establish the following infinite-dimensional result due original to Shirokov [142].

Theorem 12.5.11 (Shirokov [142]). *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and $\Psi : \mathcal{S}(\mathbb{K}_A) \to \mathcal{S}(\mathbb{K}_B)$ be arbitrary channels. The following properties are equivalent:*
1. *Equality (12.64) holds for arbitrary closed subsets $\mathcal{H} \subseteq \mathcal{S}(\mathbb{H}_A)$ and $\mathcal{K} \subseteq \mathcal{S}(\mathbb{K}_A)$ such that $H(\Phi(\rho)) < +\infty$ for all $\rho \in \mathcal{H}$ and $H(\Psi(\sigma)) < +\infty$ for all $\sigma \in \mathcal{K}$;*
2. *Inequality*
$$\chi_{\Phi\otimes\Psi}(\omega) \le \chi_\Phi(\omega_{\mathbb{H}_A}) + \chi_\Psi(\omega_{\mathbb{K}_A}) \tag{12.65}$$
holds for arbitrary state ω such that $H(\Phi(\omega_{\mathbb{H}_A})) < +\infty$ and $H(\Psi(\omega_{\mathbb{K}_A})) < +\infty$;
3. *Inequality*
$$\hat{H}_{\Phi\otimes\Psi}(\omega) \ge \hat{H}_\Phi(\omega_{\mathbb{H}_A}) + \hat{H}_\Psi(\omega_{\mathbb{K}_A}) \tag{12.66}$$
holds for arbitrary state ω such that $H(\Phi(\omega_{\mathbb{H}_A})) < +\infty$ and $H(\Psi(\omega_{\mathbb{K}_A})) < +\infty$,

where
$$\hat{H}_\Phi(\rho) = \inf_{\mu\in\mathcal{P}_{\{\rho\}}} \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu(d\rho) = \inf_{\sum_i p_i\rho_i=\rho} \sum_i p_i H(\Phi(\rho_i)).$$

Proof. (1) ⇒ (3). Let ω be an arbitrary state with finite $H(\Phi(\omega_{\mathbb{H}_A}))$ and $H(\Psi(\omega_{\mathbb{K}_A}))$. The validity of (1) implies
$$C_\chi(\Phi\otimes\Psi;\{\omega_{\mathbb{H}_A}\}\otimes\{\omega_{\mathbb{K}_A}\}) = C_\chi(\Phi;\{\omega_{\mathbb{H}_A}\}) + C_\chi(\Psi;\{\omega_{\mathbb{K}_A}\}).$$

By Lemma 12.5.10, the state $\Phi(\omega_{\mathbb{H}_A}) \otimes \Psi(\omega_{\mathbb{K}_A})$ is the output optimal average state for the $\{\omega_{\mathbb{H}_A}\} \otimes \{\omega_{\mathbb{K}_A}\}$-constrained channel $\Phi \otimes \Psi$. Noting that $\omega \in \{\omega_{\mathbb{H}_A}\} \otimes \{\omega_{\mathbb{K}_A}\}$ and applying Corollary 12.2.20, we obtain
$$\begin{aligned}\chi_\Phi(\omega_{\mathbb{H}_A}) + \chi_\Psi(\omega_{\mathbb{K}_A}) &= C_\chi(\Phi;\{\omega_{\mathbb{H}_A}\}) + C_\chi(\Psi;\{\omega_{\mathbb{K}_A}\})\\ &= C_\chi(\Phi\otimes\Psi;\{\omega_{\mathbb{H}_A}\}\otimes\{\omega_{\mathbb{K}_A}\})\\ &\ge \chi_{\Phi\otimes\Psi}(\omega) + H((\Phi\otimes\Psi)(\omega)\|\Phi(\omega_{\mathbb{H}_A})\otimes\Psi(\omega_{\mathbb{K}_A})).\end{aligned} \tag{12.67}$$

Due to the fact that

$$\begin{aligned}&H\big((\Phi\otimes\Psi)(\omega)\|\Phi(\omega_{\mathbb{H}})\otimes\Psi(\omega_{\mathbb{K}})\big)\\&\quad=H\big(\Phi(\omega_{\mathbb{H}})\big)+H\big(\Psi(\omega_{\mathbb{K}})\big)-H\big((\Phi\otimes\Psi)(\omega)\big),\end{aligned}$$

the inequality (12.67) together with the representation that

$$\chi_{\Phi}(\rho)=H\big(\Phi(\rho)\big)-\inf_{\pi\in\mathcal{P}_{\{\rho\}}}\int_{\mathcal{S}(\mathbb{H}_A)}H\big(\Phi(\rho)\big)\pi(d\rho) \tag{12.68}$$

imply (12.66).

(3) ⇒ (2). It can be derived from expression (12.68) for the Holevo χ-quantity and subadditivity of the (output) entropy.

(2) ⇒ (1). It follows from the definition of the Holevo χ-capacity (see Definition 12.1.2) and inequality (12.65) that

$$C_{\chi}(\Phi\otimes\Psi;\mathcal{H}\otimes\mathcal{K})\le C_{\chi}(\Phi;\mathcal{H})+C_{\chi}(\Psi;\mathcal{K}).$$

Since the converse inequality is obvious, there is equality here. This proves the theorem. □

The validity of inequality (12.65) for arbitrary $\omega\in\mathcal{S}(\mathbb{H}_A\otimes\mathbb{K}_A)$ seems to be substantially stronger than the equivalent properties in Theorem 12.5.11. This property is called *subadditivity* of the Holevo χ-quantity for the constrained channels Φ and Ψ. By using arguments from the proof of Theorem 12.5.11, it is easy to see that subadditivity of the Holevo χ-quantity or the channels Φ and Ψ is equivalent to validity of equality (12.64) for arbitrary subsets $\mathcal{H}\subset\mathcal{S}(\mathbb{H}_A)$ and $\mathcal{K}\subset\mathcal{S}(\mathbb{K}_A)$.

By using Proposition 12.5.8 below, it is possible to show that properties (1)–(3) in the above theorem are equivalent to subadditivity of the Holevo χ-quantity for the channels Φ and Ψ having the following property: $H(\Phi(\rho))<+\infty$ and $H(\Psi(\sigma))<+\infty$ for arbitrary finite rank states $\rho\in\mathcal{S}(\mathbb{H}_A)$ and $\sigma\in\mathcal{S}(\mathbb{K}_A)$.

Remark 12.3. By Theorem 1 in Holevo and Shirokov [79], the subadditivity of the Holevo χ-quantity for arbitrary finite-dimensional channels Φ and Ψ is equivalent to validity of inequality (12.66) for arbitrary state $\omega\in\mathcal{S}(\mathbb{H}_A\otimes\mathbb{K}_A)$, which implies additivity of the minimal output entropy

$$\inf_{\omega\in\mathcal{S}(\mathbb{H}_A\otimes\mathbb{K}_A)}H\big((\Phi\otimes\Psi)(\omega)\big)=\inf_{\rho\in\mathcal{S}(\mathbb{H}_A)}H\big(\Phi(\rho)\big)+\inf_{\sigma\in\mathcal{S}(\mathbb{K}_A)}H\big(\Psi(\sigma)\big) \tag{12.69}$$

for these channels. This follows from the inequality

$$\begin{aligned}H\big((\Phi\otimes\Psi)(\omega)\big)&\ge\hat{H}_{\Phi\otimes\Psi}(\omega)\ge\hat{H}_{\Phi}(\omega_{\mathbb{H}_A})+\hat{H}_{\Psi}(\omega_{\mathbb{K}_A})\\&\ge\inf_{\rho\in\mathcal{S}(\mathbb{H}_A)}H\big(\Phi(\rho)\big)+\inf_{\sigma\in\mathcal{S}(\mathbb{K}_A)}H\big(\Psi(\sigma)\big)\end{aligned} \tag{12.70}$$

valid for arbitrary state $\omega \in \mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$ for which inequality (12.66) holds. In contrast to this in the infinite-dimensional case, we cannot prove the above implication (without some additional assumptions). The problem consists in existence of pure states in $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$ with infinite entropies of partial traces, which can be called *superentangled*. To show this, note first that the monotonicity property of the relative entropy Linblad [108] provides the following inequality:

$$\begin{aligned} H(\omega_{\mathbb{H}_A}) + H(\omega_{\mathbb{K}_A}) - H(\omega) &= H(\omega \| \omega_{\mathbb{H}_A} \otimes \omega_{\mathbb{K}_A}) \\ &\geq H\big((\Phi \otimes \Psi)(\omega) \| \Phi(\omega_{\mathbb{H}_A}) \otimes \Psi(\omega_{\mathbb{K}_A})\big) \\ &= H\big(\Phi(\omega_{\mathbb{H}_A})\big) + H\big(\Psi(\omega_{\mathbb{K}_A})\big) - H\big((\Phi \otimes \Psi)(\omega)\big), \end{aligned}$$

which shows that $H(\omega_{\mathbb{H}_A}) = H(\omega_{\mathbb{K}_A}) < +\infty$ implies $H(\Phi(\omega_{\mathbb{H}_A})) < +\infty$ and $H(\Psi(\omega_{\mathbb{K}_A})) < +\infty$ for arbitrary pure state $\omega \in \mathcal{S}(\mathbb{H}_A \otimes \mathbb{K}_A)$ with finite output entropy $H((\Phi \otimes \Psi)(\omega))$. By this and Theorem 12.5.11, the subadditivity of the Holevo χ-quantity for arbitrary infinite-dimensional channels Φ and Ψ implies validity of inequality (12.66), and hence, validity of inequality (12.70) for all pure states ω such that $H(\omega_{\mathbb{H}_A}) = H(\omega_{\mathbb{K}_A}) < +\infty$ and $H((\Phi \otimes \Psi)(\omega)) < +\infty$. So, if we considered only such states ω in the calculation of the minimal output entropy for the channel we would obtain that it is equal to the sum of $\inf_{\rho \in \mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))$ and $\inf_{\sigma \in \mathcal{S}(\mathbb{K}_A)} H(\Psi(\sigma))$, but this additivity can be (probably) broken by taking into account superentangled states.

13 Classical capacities of memoryless channels

In a typical communication scenario, two parties (the sender Alice and the receiver Bob) aim to exchange (classical or quantum) information through a communication channel. However, the biggest hurdle in the path of efficient information transmission is the presence of noise in the communication channel. Noise distorts the information sent through the channel and the receiver often does not receive the original information that is intended. To overcome this hurdle, *quantum error correcting codes* (QECC) can be used to combat the effects of noise.

One of the most important problems of quantum information theory is to determine the capacity of noisy quantum channels. The channel capacity is defined as the optimal rate at which information may be transferred with the vanishing error in the limit of a large number of channel uses.

A classical channel has a unique capacity. However, quantum channels have several distinct capacities, depending on (1) the nature of transmitted information (classical or quantum); (2) the nature of the input states (entangled state or product states); (3) the nature of measurement done on the output (collective or individual); and (4) the presence or absence of any additional resources such as prior shared entanglement between Alice and Bob and (5) whether Alice and Bob are allowed to communicate classically with each other.

In most works on the transmission of information via a noisy quantum channel such as those treated in Hayashi [61], Ohya and Petz [121], Watrous [173], Wilde [179], Holevo [77] and references contained therein, it has usually been assumed that (i) the noisy channel acts independently and identically for each channel use and behaves as if it does not have any memory on any previous transmissions, and is therefore termed as *quantum memoryless channels*; and (ii) the Hilbert spaces are all finite- dimensional (and hence all results obtained therein are essential finite-dimensional). Quite contrary to the finite-dimensional case, the techniques required in treating infinite dimensionality of these issues are sophisticated and the mathematical challenges posed by these problems are enormous.

The purpose of this chapter is to systematically present an infinite-dimensional version of classical capacities for various types of memoryless quantum channels, including unconstrained and constrained hybrid and entanglement-assisted channels. While there are results available on this subject, the exposition presented here have never appeared in a book form before and, therefore, will be very beneficial to the novice and seasoned researcher alike.

The capacities of various types for the memory channel will be the subject of study in next three chapters.

What is a memoryless quantum channel? To describe a memoryless quantum channel mathematically, let $\mathbb{H}_{\Lambda,A}^{(n)} := \otimes_{j=1}^{N_n}(\mathbb{H}_{j,A}^{\otimes n})$ be the Hillbert space representing the quantum system of the N_n encoded quantum states (or carriers) corresponding to the

https://doi.org/10.1515/9783110788105-013

classical signals $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the system and let $\rho^{(n)}_{\Lambda,A} = (\rho^{(n)}_1, \rho^{(n)}_2, \ldots, \rho^{(n)}_{N_n}) \in \mathcal{S}(\mathbb{H}^{(n)}_{\Lambda,A})$ be the codeword of length n and size N_n be the input state on $\mathbb{H}^{(n)}_{\Lambda,A}$ encoded by the sender that represents the encoded signal of the first n signals (or carriers) for $n = 1, 2, \ldots$. The quantum channel Φ is said to be memoryless, if the first n uses the channel $\Phi^{(n)} : \mathcal{S}(\mathbb{H}^{(n)}_{\Lambda,A}) \to \mathcal{S}(\mathbb{H}^{(n)}_{\Lambda,B})$ to transmit the classical signals $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ with the encoded quantum state $\rho^{(n)}_{\Lambda,A} \in \mathcal{S}(\mathbb{H}^{(n)}_{\Lambda,A})$ and can be written as

$$\Phi^{(n)}(\rho^{(n)}_{\Lambda,A}) = \Phi^{\otimes n}(\rho^{(n)}_{\Lambda,A}), \tag{13.1}$$

where $\Phi^{\otimes n}$ is the n-folds tensor product of Φ, i. e., $\Phi^{\otimes n} = \Phi \otimes \cdots \otimes \Phi$ and $\Phi : \mathcal{S}(\mathbb{H}_{\lambda,A}) \to \mathcal{S}(\mathbb{H}_{\lambda,B})$. Here, $\mathbb{H}_{\lambda,A}$ denotes the Hilbert space representing the quantum for the codeword ρ_λ of the classical signal λ prepared and sent by Alice and $\mathbb{H}_{\lambda,B}$ denotes the Hilbert space for the codeword ρ_λ sent through the channel and received by Bob. It can be verified that $\Phi^{(n)}$ is a quantum channel as defined in Definition 5.2.1. That is, $\Phi^{(n)}$ is a completely positive and trace preserving map from $\mathcal{S}(\mathbb{H}^{\otimes n}_{\Lambda,A})$ to $\mathcal{S}(\mathbb{H}^{\otimes n}_{\Lambda,B})$.

We give the justification of (13.1) for the memoryless channel Φ as follows.

In the scenario where Alice uses a memoryless channel, $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$, to send the joint quantum states $q^{(n)} = (q_1, q_2, \ldots, q_n) \in \mathcal{S}(\mathbb{H}^{\otimes n}_A)$ to Bob. The channel Φ is invoked n times independently. The kth invocation is given by the tensor product

$$\Phi_k = \underbrace{\mathfrak{I} \otimes \cdots \otimes \mathfrak{I}}_{k-1} \otimes \Phi \otimes \underbrace{\mathfrak{I} \otimes \cdots \otimes \mathfrak{I}}_{n-k} \tag{13.2}$$

such that the overall operation becomes the concatenation

$$\Phi^{(n)} = \Phi_n \circ \Phi_{n-1} \circ \cdots \circ \Phi_2 \circ \Phi_1 = \Phi^{\otimes n}, \tag{13.3}$$

where $\mathfrak{I}$ is the identity map from $\mathcal{S}(\mathbb{H}_A)$ to $\mathcal{S}(\mathbb{H}_B)$. Hence, Bob receives the output $\Phi^{(n)}(\rho^{(n)}) = \Phi^{\otimes n}(\rho^{(n)})$ of the joint quantum state. This is the appropriate model for the situation where the effect of memory is not present or ignored.

For notational simplicity, we often suppress one or both of the subscripts Λ and A from $\mathbb{H}_{\Lambda,A}$, $\mathbb{H}^{(n)}_{\Lambda,A}$, $\rho^{(n)}_{\Lambda,A}$, etc., when there is no danger of ambiguity.

Both classical and quantum information can be transmitted through a quantum channel. A communication system is called a *classical-quantum system* or *hybrid system* if classical information/data is to be sent through the quantum channel. A question of interest in investigating a classical-quantum system is *classical capacity*, where classical capacity is roughly defined as the average number of classical bits of information that can be transmitted, with a high degree of accuracy, through each use of that channel. As is typical for information-theoretic notions, channel capacities are more formally defined in terms of asymptotic behaviors, where the limit of an increasing number of channel uses is considered.

13.1 Transmissions of classical information

This section explores various topics behind transmissions of classical information via repeated uses of a memoryless quantum channel.

13.1.1 Preparation of classical information

We briefly review the preparation (or quantization) of classical information before transmission below.

Before we can understand relevant elements of quantum information theory, it is helpful to briefly review the process involved in transmitting information via quantum channels. Similar to classical communication systems, a general quantum communication system consists of the following five major components:

- A source (the sender), which generates the (classical) message that are to be received at the destination (the receiver).
- An encoder, which turns the (classical) message generated at the source into an encoded signal before transmission.
- A channel that is the medium used to transmit the encoded signal from the source to the receiver.
- A decoder, which reconstructs/infers the original message from the signal sent through the channel; and
- A destination (the receiver), which receives the inferred message.

Prior to transmission of the classical information through the quantum channel, the sender will have to prepare and encode the classical information into quantum states. To describe preparation of the classical data $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_{|\Lambda|}\}$ (where $|\Lambda|$ denotes the cardinality or the number of elements of the set Λ) to be sent through the quantum channel Φ, let $\mathcal{C}(\Lambda)$ be the (observable) algebra of complex-valued functions $f : \Lambda \to \mathbb{C}$ and let $\mathcal{C}^*(\Lambda)$ be the topological dual of $\mathcal{C}(\Lambda)$ (i. e., $\mathcal{C}^*(\Lambda)$ is the space of bounded linear functionals $\rho : \mathcal{C}(\Lambda) \to \mathbb{C}$). In this case, the encoding channel Ξ_E is defined to be a CPTP map from $\mathcal{C}^*(\Lambda)$ to $\mathcal{S}(\mathbb{H}_A)$.

Here, we first interpret $\mathcal{C}(\Lambda)$ as an operator algebra acting on a certain finite-dimensional Hilbert space $\mathbb{H}_C$, by choosing a fixed orthonormal basis $\{|\lambda_i\rangle\}_{i=1}^{|\Lambda|}$ in $\mathbb{H}_C := \mathbb{C}^{|\Lambda|}$, and we identify the function $f \in \mathcal{C}(\Lambda)$ with the operator $\hat{f} := \sum_{i=1}^{|\Lambda|} f(\lambda_i)|\lambda_i\rangle\langle\lambda_i| \in \mathfrak{B}(\mathbb{H}_C)$. Hence, $\mathcal{C}(\Lambda)$ can be thought of as diagonal $|\Lambda| \times |\Lambda|$-matrices.

The function $f \in \mathcal{C}(\Lambda)$ is said to be an *effect* if $0 \le f(\lambda_i) \le 1$ for all $1 \le i \le |\Lambda|$. The collection of all effects in $\mathcal{C}(\Lambda)$ is denoted by $\mathcal{E}(\mathcal{C}(\Lambda))$. If $f \in \mathcal{E}(\mathcal{C}(\Lambda))$, we can physically interpret $f(\lambda_i)$ as the probability that the effect f registers the elementary event $\lambda_i \in \Lambda$. Besides, $p \in \mathcal{E}(\mathcal{C}(\Lambda))$ is a *proposition* if and only if $p(\lambda_i) \in \{0, 1\}$ for all $\lambda_i \in \Lambda$. Since $\mathcal{C}(\Lambda)$ is finite-dimensional, it is naturally isomorphic to its dual $\mathcal{C}^*(\Lambda)$ in the following sense: each linear functional $\rho \in \mathcal{C}^*(\Lambda)$ is in one-to-one corresponding to the function

$\lambda_i \mapsto \rho_{\lambda_i} := \rho(|\lambda_i\rangle\langle\lambda_i|)$, and writing shorthand $f_{\lambda_i} := f(\lambda_i)$, we have $\rho(f) = \sum_{i=1}^{|\Lambda|} f_{\lambda_i}\rho_{\lambda_i}$. As in the quantum setting, we will identify the function $\lambda \mapsto \rho_\lambda$ with the functional ρ, and use the same symbol for both. Positivity of $\rho \in \mathcal{C}^*(\Lambda)$ is equivalent to the requirement that $\rho_{\lambda_i} \geq 0, \forall\lambda_i \in \Lambda$, and the normalization becomes $1 = \rho(\mathbf{I}_{\mathcal{C}(\Lambda)}) = \rho(\sum_{i=1}^{|\Lambda|} |\lambda_i\rangle\langle\lambda_i|)$. Hence, the state $\rho \in \mathcal{C}^*(\Lambda)$ corresponds to a discrete probability distribution $\{\rho_{\lambda_i}\}_{i=1}^{|\Lambda|}$, and ρ_{λ_i} is the probability that the elementary event that λ_i occurs when the system is in the state ρ. More generally, $\sum_{i=1}^{|\Lambda|} f_{\lambda_i}\rho_{\lambda_i}$ is the probability to measure the effect f when the system is prepared in the state ρ. The pure states of the system are the Dirac measures $\{\delta_{\lambda_i}\}_{i=1}^{|\Lambda|}$, with $\delta_\lambda(|\eta\rangle\langle\eta|) := \delta_{\lambda\eta}$ being the Dirac delta function.

The prepared state $\rho \in \mathcal{C}^*(\mathbb{H})$ will be encoded by the encoding channel Ξ_E into a state in $\mathcal{S}(\mathbb{H}_A)$ and sent through the quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$. The operation of encoding channel Ξ_E is to be described in detail in the following subsection.

13.1.2 Encoding and decoding of classical information

In the following, we explore and mathematically formulate the encoding and decoding channels when information to be transmitted is classical data. In this case, the channel Φ is referred to as a hybrid or a classical-quantum channel.

When the information to be transmitted is a very long data stream, Alice will first break the stream up into pieces of some fixed length using a so-called block code. Each of these such pieces is called message. A message $w = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is said to have length n if it consists of n symbols $\lambda_i, i = 1, 2, \ldots, n$; each is chosen at random from the alphabet Λ that consists of countably infinite symbols. The procedure given by the block code encodes each message individually into a codeword, and is also i called a *block* in the context of block codes. Prior to transmitting the message, Alice will first code the message in a block into an N-component vector of quantum states $(\rho_1^{(n)}, \rho_2^{(n)}, \ldots, \rho_N^{(n)})$ referred as the codeword of size N, where $\rho_k^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n})$ for $k = 1, 2, \ldots, N$. She will then send the codeword through n-uses $\Phi^{(n)} = \Phi^{\otimes n}$ of the memoryless channel Φ one component at a time, where N is referred to as the size of the codeword. Upon receiving the output $\Phi^{\otimes n}(\rho_k^{(n)})$, $k = 1, 2, \ldots, N$. Bob, the receiver, will then decode by making measurements on $\Phi^{\otimes n}(\rho_k^{(n)})$ and try to infer the original message sent by Alice, the sender. In the context of the classical-quantum channel, the process of preparing the classical information to be sent through the channel is called *encoding* and the process of making a measurement on the quantum state received is called *decoding*. The performance and success of the overall transmission depends on the parameters of the channel and the block code. The discussions of the encoding channel Ξ_E and decoding channel Ξ_D are described in the paragraphs that follow. These models are general enough for transmission of quantum information as well.

To identity the encoder Ξ_E and decoder Ξ_E in the context of the classical-quantum channel, we let Λ be an alphabet that consists of countably infinite symbols λ, and let

$\Lambda^{\times n} = \Lambda \times \cdots \times \Lambda$ be the nth Cartesian product of the set Λ. Let $(\mathcal{S}(\mathbb{H}^{\otimes n}))^{\times N}$ be the Nth Cartesian product of the space $\mathcal{S}(\mathbb{H}_A^{\otimes n})$ below.

The details of the encoding channel Ξ_E and decoding channel Ξ_D are to be explored in the following for the memoryless hybrid/cq channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ through which Alice wants to communicate classical data to Bob using the block code.

To transmit information through n uses of the channel, the sender and the receiver have to agree on a code, i. e., an assignment of a sequence of input signals and a measurement operator on the output system to each possible message, such that the measurement operators form a valid quantum measurement, normally described by a positive operator valued measure (POVM).

The process of encoding and decoding of classical data communicated through quantum channels can be found in Hayashi [61] (see also Watrous [173]) and are described below.

We identity encoding channel Ξ_E and decoding channel below.

A. Encoder $\Xi_E : \Lambda \to \mathcal{S}(\mathbb{H}_A)$

First, we define a composite map of Λ with Φ, $W := \Phi \circ \Xi_E : \Lambda \to \mathcal{S}(\mathbb{H}_B)$, where Λ is an arbitrary set (called the input alphabet), and $\mathbb{H}_B$ is the complex Hilbert space that represents the output system B. That is, W maps classical input signals in Λ into output quantum states on $\mathbb{H}_B$. The map W is often referred to as a classical-quantum channel that transmits classical data from Λ down through the quantum channel Φ to the output system B. We denote the set of classical-quantum channels with input space Λ and output Hilbert space $\mathbb{H}_B$ by $\mathfrak{QC}(\Lambda, \mathbb{H}_B)$. We also often refer to the quantum channel Φ instead of $W = \Phi \circ \Xi_E$ as the classical-quantum channel when it is used to transmit classical data in the literature. Technically, according to Holevo and Shirokov [83], a channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is said to be classical-quantum if the image of the dual channel $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ consists of commuting operators. If all of these operators are diagonal in a fixed orthonormal basis $\{|k\rangle\}$ in $\mathbb{H}_A$, we say that the classical-quantum channel is of discrete type. In this case, it has the following representation:

$$\Phi(\rho) = \sum_{k=1}^{\dim(\mathbb{H}_A)} \langle k|\rho|k\rangle_{\mathbb{H}_A} \sigma_k,$$

where $\{\sigma_k\}$ is a collection of states in $\mathcal{S}(\mathbb{H}_B)$. Any finite-dimensional classical-quantum channel is of discrete type. An example of a classical-quantum channel, which is not of discrete type is provided by a Bosonic Gaussian classical-quantum channel (see Chapter 12 of Holevo [77] and the Appendix of Holevo–Shirokov [83]). For every channel $W \in \mathfrak{QC}(\Lambda, \mathbb{H}_B)$, we define the lifted channel $\mathfrak{W} : \Lambda \to \mathcal{S}(\mathbb{H}_\Lambda \otimes \mathbb{H}_B)$, $\mathfrak{W}(\lambda) := (|\lambda\rangle_\Lambda\langle\lambda|) \otimes W(\lambda)$. Here, $\mathbb{H}_\Lambda$ is an auxiliary Hilbert space representing the quantization of the classical data Λ and $\{|\lambda\rangle_\Lambda \mid \lambda \in \Lambda\}$ is an orthonormal basis in it. As a canonical choice, one can use $\mathbb{H}_\Lambda = L^2(\Lambda)$, the L^2-space on Λ with respect to the counting measure, and choose $|\lambda\rangle_\Lambda$ to be the characteristic function (indicator function) of the singleton $\{\lambda\}$. Let $\mathcal{P}_f(\Lambda)$ denote the set of finitely supported probability measures on Λ. We identify every $P \in \mathcal{P}_f(\Lambda)$ with the corresponding probability mass function, and hence write $P(\lambda)$ instead of $P(\{\lambda\})$ for every $\lambda \in \Lambda$. We can redefine every channel W with input alphabet Λ as a channel on the set of Dirac measures $\{\delta_\lambda \mid \lambda \in \Lambda\} \subset \mathcal{P}_f(\Lambda)$ by defining $W(\delta_\lambda) := W(\lambda)$. W then admits a natural affine extension to $\mathcal{P}_f(\Lambda)$, given by

$$W(P) := P(\lambda)W(\lambda), \quad \lambda \in \Lambda.$$

In particular, the extension of the lifted channel $\mathfrak{W}$ outputs classical-quantum states of the form

$$\mathfrak{W}(P) = \sum_{\lambda\in\Lambda} P(\lambda)(|\lambda\rangle_\Lambda\langle\lambda|) \otimes W(\lambda).$$

Note that the marginals of $\mathfrak{W}(P)$ are

$$\mathrm{tr}_B[\mathfrak{W}(P)] := \mathrm{tr}_{\mathbb{H}_B}[\mathfrak{W}(P)] = \sum_{\lambda\in\Lambda} P(\lambda)(|\lambda\rangle_\Lambda\langle\lambda|) \tag{13.4}$$

and

$$\mathrm{tr}_\Lambda[\mathfrak{W}(P)] := \mathrm{tr}_{\mathbb{H}_\Lambda}[\mathfrak{W}(P)] = \sum_{\lambda\in\Lambda} P(\lambda)W(\lambda) = W(P). \tag{13.5}$$

With a slight abuse of notation, we will also denote $\sum_{\lambda\in\Lambda} P(\lambda)(|\lambda\rangle_\Lambda\langle\lambda|)$ by P.

The n-fold i. i. d. extension of a channel $W : \Lambda \to \mathcal{S}(\mathbb{H}_A)$ is defined as $W^{\otimes n} : \Lambda^{\times n} \to \mathcal{S}(\mathbb{H}_A^{\otimes n})$,

$$W^{\otimes n}(\lambda) := W(\lambda_1) \otimes \cdots \otimes W(\lambda_n), \quad \lambda = (\lambda_1, \ldots, \lambda_n) \in \Lambda^{\times n}.$$

Given Λ, we will always choose the auxiliary Hilbert space $\mathbb{H}_{\Lambda^{\times n}}$ to be $\mathbb{H}_\Lambda^{\otimes n}$ and

$$|\lambda\rangle_{\mathbb{H}_\Lambda^{\otimes n}} := |\lambda_1\rangle_{\mathbb{H}_\Lambda} \otimes \cdots \otimes |\lambda_n\rangle_{\mathbb{H}_\Lambda},$$

for $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \Lambda^{\times N}$. With this convention, the lifted channel of $W^{\otimes n}$ is equal to $\mathfrak{W}^{\otimes n}$. Moreover, for every probability distribution $P \in \mathcal{P}_f(\Lambda)$,

$$W^{\otimes n}(P^{\otimes n}) = \big(W(P)\big)^{\otimes n} \quad \text{and} \quad \mathfrak{W}^{\otimes n}(P^{\otimes n}) = \big(\mathfrak{W}(P)\big)^{\otimes n},$$

where $P^{\otimes n} \in \mathcal{P}(\lambda)$, $P(\lambda) := P(\lambda_1)\cdots P(\lambda_n)$, $\lambda = (\lambda_1, \ldots, \lambda_n) \in \Lambda^{\times n}$, denotes the nth i. i. d. extension of P.

Let $W : \Lambda \to \mathcal{S}(\mathbb{H}_B)$ be a classical-quantum channel, as described in part (A) above. The encoding and decoding process of message transmission over the n-fold extension of the channel is described as follows. Each message $k \in \{1, 2, \dots, N\}$ (where $N = N_n$ is a positive integer depending on n) is encoded to a codeword by an encoder $\varphi_n(\cdot) : \{1, 2, \dots, N_n\} \to \Lambda^{\times n}$ defined by

$$\varphi_n(k) = (\lambda_{k,1}, \lambda_{k,2}, \dots, \lambda_{k,n}) \in \Lambda^{\times n},$$

and is mapped to

$$W^{\otimes n}(\varphi_n(k)) = W(\lambda_{k,1}) \otimes W(\lambda_{k,2}) \otimes \cdots \otimes W(\lambda_{k,n}) \in \mathcal{S}(\mathbb{H}_B^{\otimes n}).$$

The set $\{\varphi_n(k)\}_{k=1}^{N_n} \subset \Lambda^{\times n}$ is called a codebook, which is agreed upon by the sender and the receiver in advance. The quantum states $\rho_k^{(n)} \in \mathcal{S}(\mathbb{H}_{k,A}^{\otimes n})$ then can be identified with $\varphi_n(k)$ through the map

$$\Phi^{\otimes n}(\rho_k^{(n)}) = (\Phi^{\otimes n} \circ \Xi_E^{\otimes n})(\varphi_n(k)) = W^{\otimes n}(\varphi_n(k)).$$

B. Decoder $\Xi_D : \{\rho_k^{(n)}\}_{k=1}^N \to \{1, 2, \dots, N\}$

The decoding process, called the decoder, is described by a POVM $\mathbf{D}^{(n)} = \{\mathbf{D}_k^{(n)}\}_{k=1}^{N_n}$ on $\mathbb{H}_B^{\otimes n}$, where the outcomes $k = 1, 2, \dots, N_n$ indicate decoded messages. The pair $\mathcal{C}_n = (\varphi_n, \mathbf{D}_n)$ is called a code and its cardinality $|\mathcal{C}^{(n)}| = N_n$ is called the size of the code $\mathcal{C}^{(n)}$.

When a message k was sent, the probability of obtaining the outcome l is given by

$$P(l|k) = \operatorname{tr}[W^{\otimes n}(\varphi_n(k))\mathbf{D}_l^{(n)}], \quad k, l = 1, 2, \dots, N.$$

The average error probability of the code $\mathcal{C}^{(n)}$ is then given by

$$\bar{\mathbb{P}}_{\text{err}}(W^{\otimes n}, \mathcal{C}^{(n)}) = 1 - \frac{1}{N} \sum_{k=1}^{N} \operatorname{tr}[W^{\otimes n}(\varphi_n(k))\mathbf{D}_k^{(n)}],$$

which is required to vanish asymptotically for reliable communication. At the same time, the aim of classical-quantum channel coding is to make the transmission rate $\liminf_{n\to+\infty} \frac{1}{n} \log |\mathcal{C}_n|$ as large as possible. The channel capacity $C_{cq}(W)$ is defined as the supremum of achievable rates with asymptotically vanishing error probabilities, i. e.,

$$C_{cq}(W) = \sup\left\{R \mid \exists (\mathcal{C}_n)_{n=1}^{+\infty} \ni \liminf_{n\to+\infty} \frac{1}{n} \log |\mathcal{C}_n| \geq R, \lim_{n\to+\infty} P_{\text{err}}(W^{\otimes n}, \mathcal{C}_n) = 0\right\}. \quad (13.6)$$

The maximum rate (the logarithm of the number of messages divided by the number of channel uses) that can be reached by such coding schemes in the asymptotics of large n, with an asymptotically vanishing probability of erroneous decoding, is the capacity of the channel. By suppressing the encoding step in the process, $C_{cq}(W)$, we often refer $C_{cq}(W)$ to $C(\Phi)$ as the classical capacity of the memoryless quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$.

13.2 Unconstrained classical capacity

Recall from Definition 12.2.1 that the Holevo χ-capacity $C_\chi(\Phi)$ of the quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is defined as

$$C_\chi(\Phi) = \max_{\{(p_i,\rho_i)\}} \left\{ H\Big(\sum_j p_j \Phi(\rho_j)\Big) - \sum_j p_j H(\Phi(\rho_j)) \right\}, \tag{13.7}$$

where the maximum is taken over all ensembles $\mu = \{(p_j, \rho_j)\}$ of all possible input states ρ_j with distribution p_j.

When $\dim(\mathbb{H}_A) = d < +\infty$, it has been shown that the maximum in (13.7) can be achieved by using an ensemble of pure states and that in the maximization it suffices to consider ensembles of at most d^2 pure states (see Theorem 12.2.9).

The classical capacity $C(\Phi)$ of a memoryless quantum channel Φ is the maximum rate at which classical data can be sent over it error-free in the limit of many uses of the channel. Precisely, we have the following.

Definition 13.2.1. Let $\Phi^{\otimes n} : \mathcal{S}(\mathbb{H}_A^{\otimes n}) \to \mathcal{S}(\mathbb{H}_B^{\otimes n})$ be the n-use of the memoryless channel Φ.

1. The triplet $\mathfrak{C}^{(n)} = (p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)})$ is said to be a code of length n and of size N_n, where (i) $p^{(n)} = \{p_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a probability distribution, i. e., $p_j^{(n)} > 0$ for all $j = 1, 2, \ldots, N_n$ with $\sum_{j=1}^{N_n} p_j^{(n)} = 1$; (ii) $\rho^{(n)} = \{\rho_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a collection of N_n states satisfying (16.58) and (iii) $\mathbf{D}^{(n)} = \{\mathbf{D}_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a POVM on $\mathbb{H}_A^{\otimes n}$ that represents the decoding operators used by the receiver, Bob.
2. The average error probability $\bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)})$ for the code $\mathfrak{C}^{(n)}$ is defined by
$$\bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)}) = \frac{1}{N_n} \sum_{j=1}^{N_n} p_j^{(n)} \{1 - \mathrm{tr}[\Phi^{\otimes n}(\rho_j^{(n)}) \mathbf{D}_j^{(n)}]\},$$
and with an abuse of using the same notation $\bar{\mathbb{P}}_{\mathrm{err}}(\cdot)$, the average error probability over all codes $\mathfrak{C}^{(n)}$ of length n and size N_n will be denoted by $\bar{\mathbb{P}}_{\mathrm{err}}(n, N_n)$.
3. The classical capacity $C(\Phi)$ of the channel Φ is defined as the least upper bound of the rates R for which $\liminf_{n\to+\infty} \mathbb{P}_{\mathrm{err}}(n, 2^{nR}) = 0$, i. e.,
$$C(\Phi) = \inf\Big\{R > 0 \mid \liminf_{n\to+\infty} \bar{\mathbb{P}}_{\mathrm{err}}(n, 2^{nR}) = 0\Big\}. \tag{13.8}$$

Note in the above definition that the classical information/symbol λ_n is encoded into the codeword $\rho^{(n)} = (\rho_1^{(n)}, \rho_2^{(n)}, \ldots, \rho_{N_n}^{(n)})$ of size N_n and the expression

$\mathrm{tr}[\Phi^{\otimes n}(\rho_j^{(n)})\mathbf{D}_j^{(n)}]$ denotes the probability that the decoder $\mathbf{D}_j^{(n)}$ infers that the j-codeword was the state sent through the n-use of the memoryless channel by a correct measurement of the output state $\Phi^{\otimes n}(\rho_j^{(n)})$ received at its destination. Therefore, $\frac{1}{N_n}\sum_{j=1}^{N_n} p_j^{(n)}\{1-\mathrm{tr}[\Phi^{\otimes n}(\rho_j^{(n)})\mathbf{D}_j^{(n)}]\}$ denotes the averaged error probability over the codeword size and is denoted by $\bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)})$. Occasionally, we also use maximum probability of error $\mathbb{P}_{\mathrm{err}}(\mathfrak{C}^{(n)})$ defined by

$$\{1-\mathrm{tr}[\Phi^{\otimes n}(\rho_j^{(n)})\mathbf{D}_j^{(n)}]\}, \quad \forall j = 1,2,\dots,N_n,$$

to replace the average probability of error $\bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)})$ defined above.

13.2.1 Holevo–Schumacher–Westmoreland theorem

A channel coding theorem for a stationary and memoryless classical-quantum channel has been established by combining the direct part shown by Holevo [69] and Schumacher–Westmoreland [139] (now referred to as the Holevo–Schumacher–Westmoreland theorem or just the HSW theorem) with the (weak) converse part, which goes back to 1970's works by Holevo [67, 68]. This HSW theorem is undoubtedly a landmark in the history of quantum information theory. We state and give a proof using a quantum version of the Feinstein lemma for the HSW theorem for the finite-dimensional channel in this chapter. While the HSW theorem illustrated here is finite-dimensional, the quantum version of the Feinstein lemma used can and will be extended to treat infinite-dimensional memory channels in Chapters 15 and 16. The HSW theorem tells us that the asymptotic rate at which classical information can be transmitted over a quantum channel Φ per channel use is given by the maximum output Holevo quantity χ across all possible signaling ensembles. Unlike the proof presented in this section, HSW employed the random coding technique. Alternative proofs have been given by Winter [180], Ogawa and Nagaoka [118] and Hayashi and Nagaoka [62]. The proof in [118] was based on the standpoint of quantum hypothesis testing and the quantum information spectrum, though it also employed an argument similar to the classical Feinstein' lemma. In Hayashi and Nagaoka [62], the technique of quantum information spectrum was used and there were no structural assumptions imposed on the quantum channels.

In this subsection, we confine ourself to finite-dimensional input system A represented by Hilbert space $\mathbb{H}_A$ and finite-dimensional output system B represented by Hilbert space $\mathbb{H}_B$.

We state and provide a proof of the Holevo–Schumacher–Westmoreland (HSW) theorem below. The HSW theorem tells us the asymptotic rate at which classical information can be transmitted over a quantum channel Φ per channel use and is given by the maximum output Holevo quantity χ across all possible signaling ensembles.

Theorem 13.2.2 (Holevo–Schumacher–Westmoreland (HSW) theorem). *Let* $\mathbb{H}_A$ *and* $\mathbb{H}_B$ *be two finite-dimensional complex Hilbert spaces. The classical capacity of a memoryless channel* $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ *is given by the expression*

$$C(\Phi) = \lim_{n\to+\infty} \frac{1}{n} C_\chi(\Phi^{\otimes n}), \tag{13.9}$$

where $C(\Phi)$ *is defined in Definition* 13.2.1.

The (HSW) theorem will be proved by the quantum version of the Feinstein lemma below. The Feinstein lemma played an important role in the coding theorem in the classical information theory. Shannon [140] in his noisy channel coding theorem showed that information can be reliably sent over a classical channel at all rates up to the channel capacity. The first rigorous proof of this fundamental theorem was provided by Feinstein [50]. He used a packing argument to find an upper bound to the maximal number of codewords that can be sent through the channel with a low probability of error. His argument is often referred to as the Feinstein' lemma.

13.2.2 Quantum version of Feinstein's lemma

To state and prove the quantum version of Feinstein's lemma, we repeatedly use the concept of quantum typicality. The idea of a typical subspace plays an important role in the proofs of many coding theorems. Its role is analogous to that of the typical set in classical information theory (see Carver and Thomas [29]). We define quantum typicality based on Nielson and Chuang [116] and Schumacher [138] below.

Let $(X_n)_{n=1}^{+\infty}$ be a sequence of independent identically distributed (i. i. d.) random variables, each is distributed as the random variable X with distribution $p_X(x)$ for $x \in \Omega_X$, where Ω_X is the sample space of the random variable X. For each $n \in \mathbb{N}$, let $X^{\times n}$ be a finite sequence of random variables defined by the n-fold Cartesian product of X, $X^{\times n} = X \times X \times \cdots \times X$, with a corresponding finite sequence of outcomes $x^{\times n} = x_1 x_2 \ldots x_n$ ($x_i \in \Omega_X, i = 1, 2, \ldots, n$). Similarly, we define $Y^{\times n}$ and $y^{\times n}$ as a finite sequence of random variables and its corresponding outcomes.

Definition 13.2.3 (Quantum typicality). Let $\rho \in \mathcal{S}(\mathbb{H}_A)$ be a quantum state on complex Hilbert space $\mathbb{H}_A$ with the spectral decomposition,

$$\rho = \sum_{x\in\Omega_X} p_X(x)(|x\rangle_A\langle x|), \quad \text{where } |x\rangle_A\langle x| := |x\rangle_{\mathbb{H}_A}\langle x|.$$

The weakly typical subspace of the state ρ is defined as the span of all vectors such that the sample mean (Shannon) entropy $\bar{S}(x^{\times n})$ is close to the true (Shannon) entropy $S(X)$ of the distribution $p_X(x)$:

$$T_\delta^{X^{\times n}} := \operatorname{span}\{|x^{\times n}\rangle \in \mathbb{H}_A^{\otimes n} \mid |\bar{S}(x^{\times n}) - S(X)| \le \delta\}, \tag{13.10}$$

where $\bar{S}(x^{\times n}) := -\frac{1}{n}\log(p_{X^{\times n}}(x^{\times n}))$ and $S(X) = -\sum_x p_X(x)\log(p_X(x))$. The projector $\mathbf{P}_{\rho,\delta}^{(n)}$ onto the typical subspace of ρ, $T_\delta^{X^{\times n}}$, is defined as

$$\mathbf{P}_{\rho,\delta}^{(n)} := \sum_{x^{\times n}\in T_\delta^{X^{\times n}}} |x^{\times n}\rangle_{\mathbb{H}_A^{\otimes n}}\langle x^{\times n}|, \tag{13.11}$$

where $T_\delta^{X^{\times n}}$ is also referred to as the set of δ-typical sequences $(T_\delta^{X^{\times n}})_{n=1}^{+\infty}$ defined by (13.10).

The following lemma follows immediately from Definition 13.2.3.

Lemma 13.2.4. *The typical operator $\mathbf{P}_{\rho,\delta}^{(n)}$ satisfies the following three properties:*

1. $\operatorname{tr}[\mathbf{P}_{\rho,\delta}^{(n)}\rho^{\otimes n}] \geq 1-\epsilon$*;*
2. $\operatorname{tr}[\mathbf{P}_{\rho,\delta}^{(n)}] \leq 2^{n(S(X)+\delta)}$ *and*
3. $2^{-n(S(X)+\delta)}\mathbf{P}_{\rho,\delta}^{(n)} \leq \mathbf{P}_{\rho,\delta}^{(n)}\rho^{\otimes n}\mathbf{P}_{\rho,\delta}^{(n)} \leq 2^{-n(S(X)-\delta)}\mathbf{P}_{\rho,\delta}^{(n)}$.

In the following, a quantum version of Feinstein's lemma is developed and used to find an alternative proof of the direct channel coding theorem for transmission of classical information through a quantum memoryless channel. The first proof of this theorem, which states that all rates up to the Holevo χ-capacity are achievable, was provided independently by Holevo [69] and Schumacher and Westmoreland [163] and referred to as the Holevo–Schmacher–Westmoreland theorem (see Theorem 13.2.2).

The following quantum version of Feinstein's lemma, due originally to Datta and Dorlas [31], is proved using a series of lemmas in which the concept of quantum typicality is used repeatedly.

Theorem 13.2.5 (Feinstein's theorem for memoryless quantum channels). *Assume that* $\dim(\mathbb{H}_A) < +\infty$ *and* $\dim(\mathbb{H}_B) < +\infty$. *Let* $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ *be a memoryless quantum channel. Given* $\epsilon > 0$*, there exists* $n_0 \in \mathbb{N}$ *such that for all* $n \geq n_0$ *there exist at least* $N(n) \geq 2^{(\chi(\Phi)-\epsilon)}$ *product states* $\bar{\rho}_1^{(n)},\ldots,\bar{\rho}_{N(n)}^{(n)} \in \mathfrak{B}(\mathbb{H}_A^{\otimes n})$ *and positive operators* $\mathbf{D}_1^{(n)},\ldots,\mathbf{D}_{N(n)}^{(n)} \in \mathfrak{B}(\mathbb{H}_B^{\otimes n})$ *such that* $\sum_{k=1}^{N(n)}\mathbf{D}_k^{(n)} \leq \mathbf{I}_n$ *and*

$$\operatorname{tr}[\Phi^{\otimes n}(\bar{\rho}_k^{(n)})\mathbf{D}_k^{(n)}] > 1-\epsilon, \quad \textit{for each } k, \tag{13.12}$$

where $\chi(\Phi)$ *is the Holevo* χ*-quantity and* $\mathbf{I}_n$ *is the identity operator on* $\mathbb{H}_B^{\otimes n}$.

Proof. Let the maximum in (13.7) be attained at an ensemble $\{(p_j,\rho_j)\}_{j=1}^J$, where $J \leq d^2$, where $d = \dim(\mathbb{H}_A)$. That is,

$$\begin{aligned} C_\chi(\Phi) &= \max_{\mu\in\mathcal{P}(\mathcal{S}(\mathbb{H}_A))}\chi_\Phi(\mu) = \chi_\Phi(\{(p_j,\rho_j)\}_{j=1}^J) \\ &= H\Bigg(\sum_{j=1}^J p_j\Phi(\rho_j)\Bigg) - \sum_{j=1}^J p_j H(\Phi(\rho_j)). \end{aligned}$$

Denote $\sigma_j = \Phi(\rho_j)$, $\bar{\sigma} = \sum_{j=1}^{J} p_j\sigma_j = \sum_{j=1}^{J} p_j\Phi(\rho_j)$ and $\bar{\sigma}_n = \bar{\sigma}^{\otimes n}$. Choose $\delta > 0$. We will relate δ to ϵ at a later stage. The typical subspace theorem (see Lemma 13.2.4) (where we let $\mathbf{P}_{\rho,\delta}^{(n)} = \mathbf{P}_n$) ensures that there exists $n_1 \in \mathbb{N}$, such that for $n \geq n_1$, there is a typical subspace $T_{\delta,\epsilon}$ with projection $\mathbf{P}_n$, such that if $\bar{\sigma}_n$ has a spectral decomposition

$$\bar{\sigma}_n = \sum_{\vec{k}\in\mathbb{J}^n} \bar{\lambda}_{\vec{k}}^{(n)}(|\psi_{\vec{k}}^{(n)}\rangle_{\mathbb{H}_B^{\otimes n}}\langle\psi_{\vec{k}}^{(n)}|),$$

then

$$\left|\frac{1}{n}\log\lambda_{\vec{k}}^{(n)} + \bar{H}(\bar{\sigma}_n)\right| < \frac{\epsilon}{3},$$

for all $\vec{k} \in \mathbb{J}^n$ such that $|\psi_{\vec{k}}^{(n)}\rangle_{\mathbb{H}_B^{\otimes n}} \in T_{\delta,\epsilon}$ and

$$\mathrm{tr}[\mathbf{P}_n\bar{\sigma}_n] > 1 - \delta.$$

Let $\mathbf{P}_{\vec{j}}^{(n)}$ be the projection operator on the subspace of $\mathbb{H}_B^{\otimes n}$ spanned by the eigenvectors of the operator $\sigma_{\vec{j}}^{(n)} = \sigma_{j_1} \otimes \sigma_{j_2} \otimes \cdots \otimes \sigma_{j_n}$ with eigenvalues $\lambda_{\vec{j},\vec{k}}^{(n)} = \Pi_{i=1}^{n}\lambda_{j_i k_i}$ such that

$$\left|\frac{1}{n}\log\lambda_{\vec{j},\vec{k}}^{(n)} + \bar{H}\right| < \frac{\epsilon}{3}, \tag{13.13}$$

where $\bar{H}$ is the average von Neumann entropy defined by $\bar{H} = \sum_{j=1}^{J} p_j H(\sigma_j)$ with $p_j \geq 0$, $\sum_{j=1}^{J} p_j = 1$ and $H(\sigma_j) = -\sigma_j \log\sigma_j$ being the von Neumann entropy of the output state $\sigma_j \in \mathcal{S}(\mathbb{H}_B)$.

For any $\delta > 0$, there exists an $n_2 \in \mathbb{N}$ such that for all $n \geq n_2$,

$$\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{P}_{\vec{j}}^{(n)}]) > 1 - \delta^2, \tag{13.14}$$

where $\mathbb{E}[\cdots]$ denotes the expected value of $[\cdots]$ with respect to the probability measure $\mathbb{P}$ defined by $\mathbb{P}(X_i = \lambda_{jk}) = \mathbb{P}(X = \lambda_{jk}) = p_j\lambda_{jk}$ for all $i = 1, 2, \ldots, n$.

Let $J \in \mathbb{N}$, and let

$$\mathbb{J}^n = \{(j_1, j_2, \ldots, j_n) \mid j_i = 1, \ldots, J, \text{ for } i = 1, 2, \ldots, n\}.$$

Let $\sigma_j \in \mathcal{S}(\mathbb{H}_B)$ for each $j = 1, \ldots, J$ with eigenvalue λ_{jk}, where $k = 1, 2, \ldots, d$, in which $d = \dim(\mathbb{H}_B)$ is the dimension of the Hillbert space $\mathbb{H}_B$.

We need the following series of lemmas for the completion of the proof of Theorem 13.2.5.

Lemma 13.2.6. *Given a finite sequence of length $n \in \mathbb{N}$ and an n-component vector $\vec{j} = (j_1, j_2, \ldots, j_n) \in \mathbb{J}^n$, let $\mathbf{P}_{\vec{j}}^{(n)}$ be the projection operator on the subspace of $\mathbb{H}_B^{\otimes n}$ spanned* *by the eigenvectors of the operator* $\sigma_{\vec{j}}^{(n)} = \sigma_{j_1} \otimes \sigma_{j_2} \otimes \cdots \otimes \sigma_{j_n}$ *with eigenvalues* $\lambda_{\vec{j},\vec{k}}^{(n)} = \Pi_{i=1}^{n} \lambda_{j_i k_i}$ *such that*

$$\left| \frac{1}{n} \log \lambda_{\vec{j},\vec{k}}^{(n)} + \bar{H} \right| < \frac{\epsilon}{3}, \tag{13.15}$$

where $\bar{H}$ *is the average von Neumann entropy defined by* $\bar{H} = \sum_{j=1}^{J} p_j H(\sigma_j)$ *with* $p_j \geq 0$, $\sum_{j=1}^{J} p_j = 1$, *and* $H(\sigma_j) = -\sigma_j \log \sigma_j$ *being the von Neumann entropy of the output state* $\sigma_j \in \mathcal{S}(\mathbb{H}_B)$. *For any* $\delta > 0$, *there exists an* $n_2 \in \mathbb{N}$ *such that for all* $n \geq n_2$,

$$\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)} \mathbf{P}_{\vec{j}}^{(n)}]) > 1 - \delta^2, \tag{13.16}$$

where $\mathbb{E}[\cdots]$ *denotes the expected value of* $[\cdots]$ *with respect to the probability measure* $\mathbb{P}$ *defined by* $\mathbb{P}(X_i = \lambda_{jk}) = p_j \lambda_{j,k}$ *for all* $i = 1, 2, \ldots, n$.

Proof of Lemma 13.2.6. Define independent and identically distributed (i. i. d.) random variables $X_1, X_2, \ldots, X_n$ with distribution given by

$$\mathbb{P}(X = \lambda_{jk}) = p_j \lambda_{j,k}, \quad \forall j = 1, 2, \ldots n, \tag{13.17}$$

where $\lambda_{j,k}$, $k = 1, 2, \ldots, d'$, are eigenvalue of the quantum state σ_j. By the weak law of large numbers,

$$\begin{aligned} \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} \log X_i &= \mathbb{E}(\log X) = \sum_{j=1}^{J} \sum_{k=1}^{d'} p_j \lambda_{j,k} \log \lambda_{j,k} \\ &= -\sum_{j=1}^{J} p_j H(\sigma_j) = -\bar{H}. \end{aligned} \tag{13.18}$$

It follows that there exists n_2 such that for $n \geq n_2$, the typical set $T_{\delta,\epsilon}^{(n)}$ of sequences of pairs $((j_1, k_1), \ldots, (j_n, k_n))$ such that

$$\left| \frac{1}{n} \sum_{i=1}^{n} \log \lambda_{j_i,k_i} + \bar{H} \right| < \frac{\epsilon}{3}, \tag{13.19}$$

satisfies

$$\mathbb{P}(T_{\delta,\epsilon}^{(n)}) = \sum_{((j_1,k_1),\ldots,(j_n,k_n)) \in T_{\delta,\epsilon}^{(n)}} \prod_{i=1}^{n} p_{j_i} \lambda_{j_i,k_i} > 1 - \delta^2. \tag{13.20}$$

Obviously,

$$\mathbf{P}_{\vec{j}}^{(n)} \geq \sum_{\vec{k}:(\vec{j},\vec{k}) \in T_{\delta,\epsilon}^{(n)}} |\psi_{\vec{j},\vec{k}}^{(n)}\rangle_{\mathbb{H}_B^{\otimes n}} \langle \psi_{\vec{j},\vec{k}}^{(n)}|, \tag{13.21}$$

and

$$\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{P}_{\vec{j}}^{(n)}]) \geq \mathbb{P}(T_{\delta,\epsilon}^{(n)}) > 1 - \delta^2. \tag{13.22}$$

This proves the lemma. □

For each $n \in \mathbb{N}$, let $N = N(n)$ be the maximal number for which there exists states $\tilde{\rho}_1^{(n)}, \tilde{\rho}_2^{(n)}, \ldots, \tilde{\rho}_N^{(n)}$ on $\mathbb{H}_A^{\otimes n}$ of the tensor product form

$$\tilde{\rho}_k^{(n)} = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n,$$

and there exists positive operators $\mathbf{D}_1^{(n)}, \mathbf{D}_2^{(n)}, \ldots, \mathbf{D}_N^{(n)}$ on $\mathbb{H}_B^{\otimes n}$ such that, defining $\tilde{\sigma}_k^{(n)} = \Phi^{\otimes n}(\tilde{\rho}_k^{(n)})$, we have:

1. $\sum_{k=1}^N \mathbf{D}_k^{(n)} \leq \mathbf{P}_n$ and
2. $\mathrm{tr}[\tilde{\sigma}_k^{(n)}\mathbf{D}_k^{(n)}] > 1 - \epsilon$ for each k, and
3. $\mathrm{tr}[\bar{\sigma}_n \mathbf{D}_k^{(n)}] \leq 2^{-n(H(\bar{\sigma}) - \bar{H} - \frac{2}{3}\epsilon)}$ for each $k = 1, 2, \ldots, n$.

For any given $\vec{j}$, define

$$\mathbf{V}_{\vec{j}}^{(n)} = \left(\mathbf{P}_n - \sum_{k=1}^{N} \mathbf{D}_k^{(n)}\right)^{1/2} \mathbf{P}_n \mathbf{P}_{\vec{j}}^{(n)} \mathbf{P}_n \left(\mathbf{P}_n - \sum_{k=1}^{N} \mathbf{D}_k^{(n)}\right)^{1/2}. \tag{13.23}$$

Clearly, $\mathbf{V}_{\vec{j}}^{(n)} \leq \mathbf{P}_n - \sum_{k=1}^N \mathbf{D}_k^{(n)}$. We also have the following result.

Lemma 13.2.7. *Let*

$$W_n = \{\vec{j} \in \mathbb{J}^n \mid \mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{P}_{\vec{j}}^{(n)}] > 1 - \delta\}. \tag{13.24}$$

Then, for all $\vec{j} \in W_n$,

$$\mathrm{tr}[\bar{\sigma}_n \mathbf{V}_{\vec{j}}^{(n)}] \leq 2^{-n(H(\bar{\sigma}) - \bar{H} - \frac{2}{3}\epsilon)}. \tag{13.25}$$

Proof of Lemma 13.2.7. Put $\mathbf{Q}_n = \sum_{k=1}^{N(n)} \mathbf{D}_k^{(n)}$. In this case, $\mathbf{Q}_n$ and $\mathbf{P}_n$ commute, i. e., $\mathbf{Q}_n\mathbf{P}_n = \mathbf{P}_n\mathbf{Q}_n$ for each n. Using the fact that $\mathbf{P}_n\bar{\sigma}_n\mathbf{P}_n \leq 2^{-n(H(\bar{\sigma}) - \frac{1}{3}\epsilon)}$ by (13.15), we have

$$\begin{aligned}
\mathrm{tr}[\bar{\sigma}_n \mathbf{V}_{\vec{j}}^{(n)}] &= \mathrm{tr}[\bar{\sigma}_n(\mathbf{P}_n - \mathbf{Q}_n)^{1/2}\mathbf{P}_n\mathbf{P}_{\vec{j}}^{(n)}\mathbf{P}_n(\mathbf{P}_n - \mathbf{Q}_n)^{1/2}] \\
&= \mathrm{tr}[\mathbf{P}_n\bar{\sigma}_n\mathbf{P}_n(\mathbf{P}_n - \mathbf{Q}_n)^{1/2}\mathbf{P}_{\vec{j}}^{(n)}(\mathbf{P}_n - \mathbf{Q}_n)^{1/2}] \\
&\leq 2^{-n(H(\bar{\sigma}_n) - \frac{1}{3}\epsilon)}\, \mathrm{tr}[(\mathbf{P}_n - \mathbf{Q}_n)^{1/2}\mathbf{P}_{\vec{j}}^{(n)}(\mathbf{P}_n - \mathbf{Q}_n)^{1/2}] \\
&\leq 2^{-n(H(\bar{\sigma}_n) - \frac{1}{3}\epsilon)}\, \mathrm{tr}[\mathbf{P}_{\vec{j}}^{(n)}] \leq 2^{-n(H(\bar{\sigma}_n) - \bar{H} - \frac{2}{3}\epsilon)},
\end{aligned}$$

where, in the last inequality, we used the standard upper bound on the dimension of the typical subspace: $\mathrm{tr}[\mathbf{P}_{\vec{j}}^{(n)}] \le 2^{n(\bar{H}+\frac{1}{3}\epsilon)}$, which follows from (13.15). This proves the lemma. □

Since $N(n)$ is maximal, it follows that for $\vec{j} \in W_n$,

$$\mathrm{tr}[\sigma_{\mathbf{j}}^{(n)} \mathbf{V}_{\vec{j}}^{(n)}] \le 1 - 2\epsilon. \tag{13.26}$$

We now show that the set W_n has high probability.

Lemma 13.2.8. *$\mu(W_n) > 1 - \delta$, where $\mu(W_n) := \sum_{\vec{j}\in W_n} p_{\vec{j}}^{(n)}$.*

Proof. If $\vec{j} \notin W_n$, then $\mathrm{tr}[\sigma_{\vec{j}}^{(n)} \mathbf{P}_{\vec{j}}^{(n)}] \le 1 - \delta$. Hence,

$$\sum_{\vec{j}\notin W_n} p_{\vec{j}}^{(n)} \,\mathrm{tr}[\sigma_{\vec{j}}^{(n)}(\mathbf{I}_A^{(n)} - \mathbf{P}_{\vec{j}}^{(n)})] \ge \delta\mu(W_n^c), \tag{13.27}$$

where $\mathbf{I}_A^{(n)}$ is the identity operator on the tensor product Hilbert space $\mathbb{H}_A^{\otimes n}$. On the other hand, by (13.16),

$$\sum_{\vec{j}\notin W_n} p_{\vec{j}}^{(n)} \,\mathrm{tr}[\sigma_{\vec{j}}^{(n)}(\mathbf{I}_A^{(n)} - \mathbf{P}_{\vec{j}}^{(n)})] \le \mathbb{E}(\mathrm{tr}[(\mathbf{I}_A^{(n)} - \mathbf{P}_{\vec{j}}^{(n)})]) < \delta^2. \tag{13.28}$$

Hence, $\mu(W_n^c) < \frac{\delta^2}{\delta} = \delta$. This proves the lemma. □

Lemma 13.2.9. *Assume $\eta < \frac{1}{3}\epsilon$ and $\delta < \epsilon$. Then for $n \ge n_3$,*

$$\mathrm{tr}\left[\bar{\sigma}_n \sum_{k=1}^{N} \mathbf{D}_k^{(n)}\right] = \mathbb{E}\left(\mathrm{tr}\left[\sigma_{\vec{j}}^{(n)} \sum_{k=1}^{N} \mathbf{D}_k^{(n)}\right]\right) \ge \eta^2. \tag{13.29}$$

Proof of Lemma 13.2.9. Define

$$\mathbf{Q}_n' = \mathbf{P}_n - (\mathbf{P}_n - \mathbf{Q}_n)^{1/2}, \tag{13.30}$$

where $\mathbf{Q}_n = \sum_{k=1}^{N} \mathbf{D}_k^{(n)}$. It follows that

$$\begin{aligned} 1 - \epsilon &\ge \mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}(\mathbf{P}_n - \mathbf{Q}_n')\mathbf{P}_{\vec{j}}^{(n)}(\mathbf{P}_n - \mathbf{Q}_n')]\} \\ &= \mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{P}_n\mathbf{P}_{\vec{j}}^{(n)}]\} - \mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'\mathbf{P}_{\vec{j}}^{(n)}\mathbf{P}_n]\} \\ &\quad - \mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{P}_n\mathbf{P}_{\vec{j}}^{(n)}\mathbf{Q}_n']\} + \mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'\mathbf{P}_{\vec{j}}^{(n)}\mathbf{Q}_n']\} \end{aligned} \tag{13.31}$$

Since the last term is positive, by Lemma 13.2.9 we have

$$\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'\mathbf{P}_{\vec{j}}^{(n)}\mathbf{P}_n] + \mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{P}_n\mathbf{P}_{\vec{j}}^{(n)}\mathbf{Q}_n']\} > \epsilon - \eta > 2\eta. \tag{13.32}$$

On the other hand, using the Cauchy–Schwarz inequality (1.2) for each term, we have

$$\begin{aligned}&\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'\mathbf{P}_{\vec{j}}^{(n)}\mathbf{P}_n] + \mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{P}_n\mathbf{P}_{\vec{j}}^{(n)}\mathbf{Q}_n']\}\\ &\quad\le 2\{\mathbb{E}(\mathrm{tr}[\mathbf{Q}_n'\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'])\}^{1/2}\{\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{P}_n\mathbf{P}_{\vec{j}}^{(n)}\mathbf{P}_n])\}^{1/2}\\ &\quad\le 2\{\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'^2])\}^{1/2}.\end{aligned} \tag{13.33}$$

Thus,

$$\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'^2]\} \ge \eta^2 \tag{13.34}$$

To complete the proof of this lemma, we now claim that

$$\mathbf{Q}_n \ge \mathbf{Q}_n'^2. \tag{13.35}$$

Indeed, this follows on the domain of $\mathbf{P}_n$ from the inequality $1-(1-x)^2 \ge x^2$ for $0 \le x \le 1$ This proves the lemma. □

To complete the proof of Theorem 13.2.5, we now have by assumption

$$\mathrm{tr}[\bar{\sigma}_n\mathbf{D}_k^{(n)}] \le 2^{-n(H(\bar{\sigma})-\bar{H}-\frac{2}{3}\epsilon)} \tag{13.36}$$

for all $k = 1,\dots,N(n)$. On the other hand, choosing $\eta < \frac{1}{3}\epsilon$ and $\delta < \frac{1}{3}\eta$, we have by Lemma 13.2.9,

$$\mathrm{tr}\left[\bar{\sigma}_n\sum_{k=1}^{N}\mathbf{D}_k^{(n)}\right] \ge \eta^2,$$

provided that $n \ge n_3$. It follows from the definition of $N(n)$ that

$$N(n) \ge \eta^2 2^{n(H(\bar{\sigma})-\bar{H}-\frac{2}{3}\epsilon)} \ge 2^{n(H(\bar{\sigma})-\bar{H}-\epsilon)}$$

for $n \ge n_3$ and $n \ge -\frac{6}{\epsilon}\log\eta$. This proves the theorem. □

Note that the quantum in Feinstein's lemma presented in this subsection, based on the results obtained in Datta and Dorlas [32], can be extended explicitly to a class of quantum channels with memory (see Chapter 14). This allows us to obtain a rigorous lower bound to the maximum achievable rate of transmission for this class of channels, for the case of product state inputs. The generalized quantum in Feinstein's lemma and the direct coding theorem for these channels with memory are given in Chapter 14.

The coding theorems for classical-quantum channel have also been extended by Winter [180, 181] and also by Mosonyi and Ogawa [115], who determine the exact strong converse exponent for classical-quantum channels.

13.3 Classical capacity of constrained channels

In this subsection, we consider the linearly constrained infinite-dimensional memoryless channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ and explore the formula of classical capacity of such a channel.

Let $\mathbf{H}$ be an $\mathfrak{H}$-operator on the input system $\mathbb{H}_A$ and consider the linearly constrained set $\mathcal{A} = \mathcal{K}_{\mathbf{H}}(E) \subset \mathcal{S}(\mathbb{H}_A)$ described by

$$\mathcal{K}_{\mathbf{H}}(E) := \{\rho \in \mathcal{S}(\mathbb{H}_A) \mid \mathrm{tr}[\rho\mathbf{H}] \leq E\}, \quad E \geq 0. \tag{13.37}$$

For arbitrary state $\rho \in \mathcal{S}(\mathbb{H}_A)$ with spectral decomposition $\rho = \sum_i \lambda_i |\phi_i\rangle_{\mathbb{H}_A}\langle\phi_i|$, we define

$$\mathrm{tr}[\rho\mathbf{H}] := \sum_i \lambda_i \|\sqrt{\mathbf{H}}|\phi_i\rangle\|^2_{\mathbb{H}_A} \leq +\infty.$$

The following constraint on the memoryless channel Φ is imposed:

$$\sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} H(\Phi(\rho)) < +\infty, \tag{13.38}$$

where E is a positive constant.

As noted earlier, the n-use of a memoryless quantum channel can be written as $\Phi^{(n)} = \Phi^{\otimes n}$, where $\Phi^{\otimes n}$ is the tensor product channel on the Hilbert space $\mathbb{H}_A^{\otimes n}$. In this case, the observable $\mathbf{H}^{(n)}$ on $\mathbb{H}_A^{\otimes n}$ corresponding to $\mathbf{H}$ on $\mathbb{H}_A$ can be defined as

$$\mathbf{H}^{(n)} = \mathbf{H} \otimes \mathbf{I}_A \otimes \cdots \otimes \mathbf{I}_A + \cdots + \mathbf{I}_A \otimes \cdots \otimes \mathbf{I}_A \otimes \mathbf{H}, \tag{13.39}$$

where $\mathbf{I}_A$ is the identity operator on the input system $\mathbb{H}_A$. We want the input state $\rho^{(n)}$ on the tensor product space $\mathbb{H}_A^{\otimes n}$ to satisfy the additive constraint

$$\mathrm{tr}[\rho^{(n)}\mathbf{H}^{(n)}] \leq nE. \tag{13.40}$$

Lemma 13.3.1. *Let* $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ *be a memoryless channel and let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on* $\mathbb{H}_A$*. If* Φ *satisfies* (13.38)*, then* $\Phi^{\otimes n}$ *satisfies the following relation:*

$$\sup_{\rho^{(n)}\in\mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} H(\Phi^{\otimes n}(\rho^{(n)})) < +\infty, \tag{13.41}$$

where

$$\mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE) = \{\rho^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n}) \mid \mathrm{tr}[\rho^{(n)}\mathbf{H}^{(n)}] \leq nE\}. \tag{13.42}$$

Proof. The subadditivity of von Neumann entropy with respect to the tensor product of channels yields

$$H(\Phi^{\otimes n}(\rho^{(n)})) \leq \sum_{k=1}^{n} H(\Phi(\rho_k^{(n)})),$$

where $\rho_k^{(n)}$ is the kth partial state of $\rho^{(n)}$, i. e., $\rho_k^{(n)}$ $(k = 1, 2, \ldots, n)$ is the partial trace of $\rho^{(n)}$ taken over the tensor product space of

$$\underbrace{\mathbb{H}_A \otimes \cdots \otimes \mathbb{H}_A}_{k-1 \text{ factors}} \otimes \underbrace{\mathbb{H}_A \otimes \cdots \otimes \mathbb{H}_A}_{n-k \text{ factors}}.$$

By concavity of the entropy,

$$\sum_{k=1}^{n} H(\Phi(\rho_k^{(n)})) \leq nH(\Phi(\bar{\rho}^{(n)})),$$

where $\bar{\rho}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} \rho_k^{(n)}$. The inequality (13.40) can be rewritten as

$$\frac{1}{n}\sum_{k=1}^{n} \mathrm{tr}[\rho_k^{(n)}\mathbf{H}^{(\mathbf{n})}] = \mathrm{tr}[\bar{\rho}^{(n)}\mathbf{H}^{(n)}] \leq E,$$

which implies that

$$\sup_{\rho^{(n)} \in \mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} H(\Phi^{\otimes n}(\rho^{(n)})) \leq n \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} H(\Phi(\rho)).$$

This proves the lemma. □

For the infinite-dimensional memoryless channel Φ satisfying the constraint (13.41), the code, error probability and classical capacity for the channel is defied below.

Definition 13.3.2. For each $n \in \mathbb{N}$, and the n-use $\Phi^{\otimes n}$ of memoryless channel Φ : $\mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ satisfying (13.41), we define the following terminologies:

1. The triplet $\mathfrak{C}^{(n)} = (p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)})$ is said to be a code of length n and of size N_n, where (i) $p^{(n)} = \{p_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a probability distribution, i. e., $p_j^{(n)} > 0$ for each $j = 1, 2, \ldots, N_n$ with $\sum_{j=1}^{N_n} p_j^{(n)} = 1$; (ii) $\rho^{(n)} = \{\rho_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a collection of N_n states satisfying (13.40) and (iii) $\mathbf{D}^{(n)} = \{\mathbf{D}_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a collection of POMV on $\mathbb{H}_A^{\otimes n}$.
2. The maximum error probability for the code $\mathfrak{C}^{(n)}$ is defined as

$$\mathbb{P}_{\mathrm{err}}(\mathfrak{C}^{(n)}) = \max_{j=1,2,\ldots,N_n} \{1 - \mathrm{tr}[\Phi^{\otimes n}(\rho_j^{(n)})\mathbf{D}_j^{(n)}]\}.$$

With an abuse of the notation, the maximum error probability over all codes $\mathfrak{C}^{(n)}$ of length n and size N_n is denoted by $\mathbb{P}_{\mathrm{err}}(n, N_n)$.

3. The energy constrained classical capacity $C(\Phi; \mathbf{H}, E)$ is defined as the least upper bound of the rates R for which $\liminf_{n\to+\infty} \mathbb{P}_{\text{err}}(n, 2^{nR}) = 0$. That is,

$$C(\Phi; \mathbf{H}, E) = \inf\Big\{R > 0 \mid \liminf_{n\to+\infty} \mathbb{P}_{\text{err}}(n, 2^{nR}) = 0\Big\}. \tag{13.43}$$

If a probability distribution $p^{(n)} = \{p_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ on the input codewords $\rho^{(n)} = \{\rho_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is given, then using the transition probability

$$p(j|k) = \operatorname{tr}[\Phi^{\otimes n}(\rho_k^{(n)})\mathbf{D}_j^{(n)}]$$

we can find the joint distribution of input and output, compute the quantum information $I_m(p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)})$ and define the quantity

$$C_\chi^{(n)}(\Phi) = \sup_{(p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)}) \in \mathfrak{C}^{(n)}} I_m(p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)}), \tag{13.44}$$

where

$$I_m(p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)}) = H\left(\sum_{k=1}^{N_n} p_k^{(n)} \Phi^{\otimes n}(\rho_k^{(n)})\right) - \sum_{k=1}^{N_n} p_k^{(n)} H(\Phi^{\otimes n}(\rho_k^{(n)})). \tag{13.45}$$

By the quantum entropy bound (13.44), we have

$$I_m(p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)}) \le C_\chi^{(n)}(\Phi), \quad \forall (p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)}) \in \mathfrak{C}^{(n)}. \tag{13.46}$$

We have the result due originally to Holevo [72, 73] and [78] that provides a computational formula for the classical capacity $C(\Phi; \mathbf{H}, E)$ in constrained infinite-dimensional quantum systems.

Proposition 13.3.3. *Assume that the channel Φ satisfies* (13.38). *Then the classical capacity of this channel under the constraint* (16.58) *is finite and equals to*

$$\begin{aligned} C(\Phi; \mathbf{H}, E) &= \lim_{n\to+\infty} \frac{1}{n} \sup_{(p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)}) \in \mathfrak{C}^{(n)}} I_m(p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)}) \\ &= \lim_{n\to+\infty} \frac{1}{n} C_\chi^{(n)}(\Phi). \end{aligned} \tag{13.47}$$

Proof. The first equality of (13.47) follows from Definition 13.3.2.

We now prove the second equality of (13.47), i. e.,

$$C(\Phi; \mathbf{H}, E) = \lim_{n\to+\infty} \frac{1}{n} C_\chi^{(n)}(\Phi).$$

We first note that the inequality $C(\Phi; \mathbf{H}, E) \le \lim_{n\to+\infty} \frac{1}{n} C_\chi^{(n)}(\Phi)$ follows from (13.46). Therefore, we only need to prove the inequality

$$C(\Phi; \mathbf{H}, E) \geq \lim_{n\to+\infty} \frac{1}{n} C_\chi^{(n)}(\Phi).$$

Assume that the transmission rate $R < C(\Phi; \mathbf{H}, E)$, then we can choose n_0, probability distribution $p^{(n_0)} = \{p_k^{(n_0)} \mid k = 1, 2, \dots, N_{n_0}\}$ and collection of states $\rho^{(n_0)} = \{\rho_k^{(n_0)} \mid k = 1, 2, \dots, N_{n_0}\}$ on $\mathbb{H}_A^{\otimes n_0}$ such that $(p^{(n_0)}, \rho^{(n_0)}, \mathbf{D}^{(n_0)}) \in \mathfrak{C}^{(n_0)}$ and

$$n_0 R < H\left(\sum_{k=1}^{N_{n_0}} p_k^{(n_0)} \Phi^{\otimes n_0}(\rho_k^{(n_0)})\right) - \sum_{k=1}^{N_{n_0}} p_k^{(n_0)} H(\Phi^{\otimes n_0}(\rho_k^{(n_0)})). \tag{13.48}$$

Consider the channel $\tilde{\Phi}$ on $\mathcal{S}(H_A^{\otimes n_0})$ given by the formula

$$\tilde{\Phi}(\rho^{(n_0)}) = \sum_k \Phi^{\otimes n_0}(\rho_k^{(n_0)}) \langle e_k | \rho^{(n_0)} | e_k \rangle, \tag{13.49}$$

and define the constraint function for this channel as $f(k) = \mathrm{tr}[\rho_k^{(n_0)} \mathbf{H}^{(n_0)}]$. The condition (13.41) implies

$$\sup_\pi H\left(\sum_{\lambda\in\Lambda} p_\lambda \Phi^{\otimes n_0}(\rho_\lambda)\right) < \infty, \tag{13.50}$$

where the supremum is over the probability distributions $p = \{p_\lambda, \lambda \in \Lambda\}$, satisfying

$$\sum_{\lambda\in\Lambda} p_\lambda f(\lambda) \leq n_0 E. \tag{13.51}$$

□

13.4 Entanglement-assisted classical capacity

The purpose of this section is to explore finite-dimensional and constrained infinite-dimensional entanglement-assisted classical capacities of the memoryless channel Φ.

Roughly speaking, the entanglement-assisted classical capacity $C_{\mathrm{ea}}(\Phi)$ is defined as the maximum amount of classical information that can be reliably transmitted over the quantum channel with the help of unlimited prior pure entanglement shared between sender and receiver.

What are the advantages of sharing pure entanglement between the sender and the receiver in transmitting classical information via a quantum channel? Entanglement is a useful resource in information transmission via a quantum channel. It can be used to enhance the performance of quantum error correcting codes, and enhance quantum channel capacities if shared between the sender and receiver prior to communication, by encoding information into entangled states when making successive uses of the channel. Precisely, one of the advantages for sharing prior entanglement between sender and receiver is that it can exactly double the classical capacity of a noiseless quantum channel and it can increase the classical capacity of some noisy quantum channels by an arbitrarily large constant factor depending on the channel, relative to the best known classical capacity achievable without entanglement. In particular, the enhancement factor is greatest for very noisy channels, with positive classical capacity but zero quantum capacity.

This section explores concepts of entanglement-assisted classical capacity for finite-dimensional and constrained infinite-dimensional memoryless channels. The formula for the entanglement-assisted capacity of a noisy quantum channel can be expressed as the maximum of mutual information over input states. The finite-dimensional case was first obtained by Bennett–Shor–Smolin–Thapliya [8] and [9]. This section presents an alternative and simpler proof for the BSST theorem given originally by Holevo [71] and [72]. A formula due originally to Holevo and Shirokov [83] for computing the entanglement-assisted classical capacity for the linearly constrained infinite-dimensional channel is also presented. The main tools for obtaining this formula are the coding theorem for classical-quantum constrained channels and a finite-dimensional approximation of the input density operators for entanglement-assisted capacity. We also give sufficient conditions under which supremum in the capacity formulas can be achieved.

To facilitate a mathematical discussion of entanglement-assisted classical capacity for a memoryless channel, we consider the following protocol of classical information transmission through quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$, (where $\mathbb{H}_{A'}$ is another quantum system, which may be different from $\mathbb{H}_A$ or $\mathbb{H}_B$ but is accessible by both the sender and receiver for sharing entangled information). Here, the systems A (possessed by Alice) and B (possessed by Bob) are represented by Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, respectively, where $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$. Suppose Alice has a set of messages, labeled by the elements of the set $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{M_n}\}$, which she would like to communicate via the quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ to Bob, exploiting this shared entanglement. In the entanglement-assisted communication, Alice (quantum system A) and Bob (quantum system B) share indefinitely many copies of an entangled (pure) state ω_{AB}. Alice first chooses the classical message λ at random from Λ_n with probability p_λ and then sends part of this shared state $\omega_{AB}^\lambda = (\mathfrak{E}_A^\lambda \otimes \mathfrak{I}_B)(\omega_{AB})$, through the channel Φ to B. Here, $\mathfrak{E}_A^\lambda$ are encoding channels depending on the signal λ. Thus, B receives states $(\Phi \otimes \mathfrak{I}_B)(\omega_{AB}^\lambda)$ with probabilities p_λ and aims to extract maximum information about λ by doing measurements on these states. To enable block encoding, this procedure should be applied to the channel $\Phi^{(n)}$. Then signal states $\omega_{AB}^{\lambda(n)} := (\omega_{AB}^\lambda)^{(n)}$ transmitted through the channel $\Phi^{(n)} \otimes \mathfrak{I}_B^{\otimes n}$ have a special form

$$\omega_{AB}^{\lambda(n)} := (\omega_\lambda^1, \omega_\lambda^2, \ldots, \omega_\lambda^n) = (\mathfrak{E}_A^{\lambda(n)} \otimes \mathfrak{I}_B^{(n)})(\omega_{AB}^{(n)}), \tag{13.52}$$

where the encoded codewords $\omega_{AB}^{(n)}$ is the pure entangled state for n copies of the system AB and $\lambda \mapsto \mathfrak{E}_A^{\lambda(n)}$ are encodings of n copies of system A.

Note that the codewords $\omega_{AB}^{\lambda(n)}$ are states shared between Alice and Bob. Alice then sends her part of these shared states to Bob through n subsequent uses of the quantum channel Φ. Hence, Bob's final state corresponding to Alice's classical message λ is

$$\sigma_{AB}^{\lambda(n)} := (\Phi^{(n)} \otimes \mathfrak{I}_B^{\otimes n})(\omega_{AB}^{\lambda(n)}).$$

In order to infer the message that Alice communicated to him, Bob makes a measurement on the received state $\sigma_{AB}^{\lambda(n)}$, the measurement being described by POVM elements $\mathbf{D}_{AB}^{(n)} = \mathbf{D}_{AB}^{\lambda(n)}$, $\lambda \in \Lambda_n$, with $\mathbf{D}_{AB}^{\lambda(n)}$ being a positive operator acting on $\mathbb{H}_{AB}^{\otimes n}$, such that

$$\sum_{\lambda \in \Lambda_n} \mathbf{D}_{AB}^{\lambda(n)} \leq \mathbf{I}_{AB}^{\otimes n},$$

and $\mathbf{I}_{AB}$ denoting the identity operator acting in $\mathbb{H}_{AB}$. Defining

$$\mathbf{D}_{AB}^{0(n)} := \mathbf{I}_{AB}^{\otimes n} - \sum_{\lambda \in \Lambda_n} \mathbf{D}_{AB}^{\lambda(n)}$$

yields a resolution of identity in $H_{AB}^{\otimes n}$. Hence, $\{\mathbf{D}_{AB}^{\lambda(n)}\}_{\lambda \in \Lambda_n \cup \{0\}}$ defines a POVM. An output $\beta \in \Lambda_n$ of a measurement described by this POVM, would lead Bob to conclude that the codeword was $\omega_{AB}^{\beta(n)}$, whereas the output 0 is interpreted as a failure of any inference. The encoding and decoding operations, employed to achieve reliable transmission of information by means of this protocol, together define a quantum code $\mathfrak{C}^{(n)}$ (of length n and size N_n), which is given by the triplet $\mathfrak{C}^{(n)} := (p^{(n)}, \omega_{AB}^{\lambda(n)}, \mathbf{D}_{AB}^{(n)})$.

13.4.1 Finite-dimensional channels

The presentation of this subsection is based on results obtained in Holevo and Shirokov [83], Holevo [71, 72] and Bennett–Shor–Smolin–Thapliya [9].

Consider the Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$, where $\dim(\mathbb{H}_A) < +\infty$ and $\dim(\mathbb{H}_B) < +\infty$. In this case, the maximum over measurements of system B can be evaluated using the coding theorem for the classical capacity (see HSW theorem (13.2.2)). First, we consider the one-shot entanglement-assisted classical capacity

$$C_{\mathrm{ea}}^{(1)}(\Phi) = \max_{(p_\lambda, \omega_{AB}^\lambda, \mathbf{D}_{AB}^\lambda)} \left\{ H\left(\sum_{\lambda \in \Lambda} p_\lambda (\Phi \otimes \mathfrak{I}_B)(\omega_{AB}^\lambda) \right) - \sum_{\lambda \in \Lambda} p_\lambda H((\Phi \otimes \mathfrak{I}_B)(\omega_{AB}^\lambda)) \right\}, \tag{13.53}$$

where the maximum is taken among all $p_\lambda > 0$ with $\sum_{\lambda \in \Lambda} p_\lambda = 1$, all shared entangled state ω_{AB}^λ corresponding to of classical message $\lambda \in \Lambda_n$, and $H(\cdot)$ is the von Neumann entropy function. This quantity is obtained by the maximization of the quantum mutual information for single channel use, which yields one-use (or one-shot) entanglement-assisted classical capacity $C_{\mathrm{ea}}^{(1)}(\Phi)$ that can easily be written as

$$C_{\mathrm{ea}}^{(1)}(\Phi) = \max_{\rho \in \mathcal{S}(\mathbb{H}_A)} \{H(\rho) - H(\Phi(\rho)) - H((\Phi \otimes \mathfrak{I}_B)(|\omega\rangle_{AB}\langle\omega|))\}, \tag{13.54}$$

where $|\omega\rangle_{AB}\langle\omega| \in \mathcal{S}(\mathbb{H}_{AB})$ ($\mathbb{H}_{AB} = \mathbb{H}_A \otimes \mathbb{H}_B$) is the maximally entangled pure state shared between the sender and the receiver, which is also a purification of the input state $\rho \in \mathcal{S}(\mathbb{H}_A)$. In this case, part of the purification, namely $\mathrm{tr}_B[|\omega\rangle_{AB}\langle\omega|]$, is transmitted through the channel Φ while the other part $\mathrm{tr}_A[|\omega\rangle_{AB}\langle\omega]$ is sent through the identity channel $\mathfrak{I}_B$ (this corresponds to the portion of the entangled state that the receiver holds at the start of the protocol).

Using the memoryless channel Φ n times and allowing entangled measurement on system B, one gets

$$C_{\mathrm{ea}}^{(n)}(\Phi) = C_{\mathrm{ea}}^{(1)}(\Phi^{\otimes n}). \tag{13.55}$$

The full entanglement-assisted classical capacity $C_{\mathrm{ea}}(\Phi)$ for the memoryless channel Φ can then be written as

$$C_{\mathrm{ea}}(\Phi) = \lim_{n\to+\infty} \frac{1}{n} C_{\mathrm{ea}}^{(n)}(\Phi) = \lim_{n\to+\infty} \frac{1}{n} C_{\mathrm{ea}}^{(1)}(\Phi^{\otimes n}). \tag{13.56}$$

The following theorem for finite-dimensional spaces $\mathbb{H}_A$ and $\mathbb{H}_B$ is due originally to Bennett–Shor–Smolin–Thapliya [8] and [9]. However, a simplified version of the proof by Holevo [71] is provided below.

Theorem 13.4.1 (The BSST theorem, [8, 9]). *Assume that* $\dim(\mathbb{H}_A) = \dim(\mathbb{H}_B) < +\infty$. *Then*

$$C_{\mathrm{ea}}(\Phi) = \max_{\omega_A} I_m(\omega_A; \Phi), \tag{13.57}$$

where $\omega_A = \mathrm{tr}_B[\omega_{AB}]$ *and*

$$I_m(\omega_A; \Phi) = H(\omega_A) + H(\Phi(\omega_A)) - H(\omega_A; \Phi) \tag{13.58}$$

is the quantum mutual information with $H(\omega_A; \Phi)$ *being the entropy exchange defined by* $H((\Phi \otimes \mathfrak{I}_B)(|\omega\rangle_{AB}\langle\omega|))$.

The proof of Theorem 13.4.1 follows from the following two lemmas.

Lemma 13.4.2. *Assume that all conditions of Theorem* 13.4.1 *hold. Then*

$$C_{\mathrm{ea}}(\Phi) \geq \max_{\omega_A} I_m(\omega_A; \Phi). \tag{13.59}$$

Proof. To prove the lemma, we first claim that

$$C_{\mathrm{ea}}^{(1)}(\Phi^{\otimes n}) \geq I_m\left(\frac{\mathbf{P}}{\dim(\mathbf{P})}; \Phi^{(n)}\right) \tag{13.60}$$

for arbitrary projection $\mathbf{P}$ in $\mathbb{H}_A^{\otimes n}$.

To prove this claim, we let $\mathbf{P} = \sum_{k=1}^{m} |e_k\rangle_A\langle e_k|$ on $\mathbb{H}_A$, where $\{e_k\}_{k=1}^{m}$ is an orthonormal system, where $m = \dim(\mathbf{P})$. Define unitary operators $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}_{\alpha\beta}$ on $\mathbb{H}_A$ by

$$\mathbf{V}|e_k\rangle_A = \exp\left(\frac{2\pi i k}{m}\right)|e_k\rangle_A; \quad \mathbf{U}|e_k\rangle_A = |e_{k+1(\mathrm{mod}\, m)}\rangle_A, \quad k = 1, 2, \dots, m;$$
$$\mathbf{W}_{\alpha\beta} = \mathbf{U}^{\alpha}\mathbf{V}^{\beta}; \quad \alpha, \beta = 1, 2, \dots, m$$

on the subspace of $\mathbb{H}_A$ generated by $\{e_k\}_{k=1}^{m}$, and as the identity operator on its orthogonal complement in $\mathbb{H}_A$. Let

$$\psi_{AB} = \frac{1}{\sqrt{m}} \sum_{k=1}^{m} |e_k\rangle_A \otimes |e_k\rangle_B.$$

Then it is easy to show that:
(i) $\{(\mathbf{W}_{\alpha\beta} \otimes \mathbf{I}_B)|\psi_{AB}\rangle; \alpha, \beta = 1, \dots, m\}$ is an orthonormal system in $\mathbb{H}_{AB} := \mathbb{H}_A \otimes \mathbb{H}_B$; in particular, if $m = \dim(\mathbb{H}_A)$, it is a basis;
(ii) $\sum_{\alpha,\beta=1}^{m} (\mathbf{W}_{\alpha\beta} \otimes \mathbf{I}_B)|\psi_{AB}\rangle\langle\psi_{AB}|(\mathbf{W}_{\alpha\beta} \otimes \mathbf{I}_B)^* = \frac{\mathbf{P}}{m} \otimes \frac{\mathbf{P}}{m}$.

Without loss of generality, we can label the classical signal $\lambda \in \Lambda$ to be transmitted as $\lambda = (\alpha, \beta)$ with equal probabilities $1/m^2$. For the entangled state $|\psi_{AB}\rangle\langle\psi_{AB}|$, we assume the unitary encodings $\mathfrak{E}_A^{\lambda}(\omega_{AB}) = \mathbf{W}_{\alpha\beta}(\omega_{AB})\mathbf{W}_{\alpha\beta}^*$. Then we have

$$C_{\mathrm{ea}}^{(1)}(\Phi^{\otimes n}) \geq H\left(\frac{1}{m^2}\sum_{\alpha\beta}(\Phi \otimes \mathfrak{I}_B)(\omega_{AB}^{\alpha\beta})\right) - \frac{1}{m^2}\sum_{\alpha\beta} H((\Phi \otimes \mathfrak{I}_B)(\omega_{AB}^{\alpha\beta})),$$

where $\omega_{AB}^{\alpha\beta} = (\mathbf{W}_{\alpha\beta} \otimes \mathbf{I}_B)|\psi_{AB}\rangle\langle\psi_{AB}|(\mathbf{W}_{\alpha\beta} \otimes \mathbf{I}_B)^*$. Then, by the property (ii), the first term on the right-hand side of the above inequality equals

$$H\left((\Phi \otimes \mathfrak{I}_B)\left(\frac{\mathbf{P}}{m} \otimes \frac{\mathbf{P}}{m}\right)\right) = H\left(\frac{\mathbf{P}}{m}\right) + H\left(\Phi\left(\frac{\mathbf{P}}{m}\right)\right)$$

Since $\omega_{AB}^{\alpha\beta}$ is a purification of $\frac{\mathbf{P}}{m}$ in $\mathbb{H}_B$, the entropies in the $H(\frac{\mathbf{P}}{m}; \Phi)$ in second term are all equal to $H(\frac{\mathbf{P}}{m}; \Phi)$. By the expression for quantum mutual information (13.60), this proves

$$C_{\mathrm{ea}}^{(1)}(\Phi^{\otimes n}) \geq I_m\left(\frac{\mathbf{P}}{\dim(\mathbf{P})}; \Phi^{\otimes n}\right) \tag{13.61}$$

for arbitrary projection $\mathbf{P}$ in $\mathbb{H}_A^{\otimes n}$. This proves the claim.

For future use, we comment without providing a proof that the last term in the quantum mutual information (13.60), the entropy exchange $H(\omega_A; \Phi)$, is equal to the final environment entropy $H(\Phi_E(\omega_A))$, where Φ_E is a channel from the system space $\mathbb{H}_A$ to the environment space $\mathbb{H}_E$. Now let $\rho_A = \rho$ be an arbitrary state in $\mathbb{H}_A$, and let $\mathbf{P}^{n,\delta}$, be the typical projection of the state $\rho^{\otimes n}$ in $\mathbb{H}_A^{\otimes n}$. It was suggested that for the arbitrary channel Ψ from $\mathbb{H}_A$ to possibly other Hilbert space $\tilde{\mathbb{H}}$,

$$\lim_{\delta\to 0}\lim_{n\to+\infty}\frac{1}{n}H\left(\Psi^{\otimes n}\left(\frac{\mathbf{P}^{n,\delta}}{\dim(\mathbf{P}^{n,\delta})}\right)\right)=H(\Psi(\rho)),$$

which would imply by the expressions for the mutual information and the entropy exchange that

$$\lim_{\delta\to 0}\lim_{n\to+\infty}\frac{1}{n}I_m\left(\frac{\mathbf{P}^{n,\delta}}{\dim(\mathbf{P}^{n,\delta})};\Phi^{\otimes n}\right)=I_m(\rho;\Phi), \tag{13.62}$$

and hence, by (13.60), the required inequality (13.59).

We shall prove (13.62) with $\mathbf{P}^{n,\delta}$ being the *strongly typical projection* of the state $\rho^{\otimes n}$.

Let us fix small positive δ, and let γ_j be the eigenvalues, $|e_j\rangle_A$ the eigenvectors of the density operator ρ. Then the eigenvalues and eigenvectors of $\rho^{\otimes n}$ are $\gamma_{\vec{j}}=\gamma_{j_1}\cdots\gamma_{j_n}$ and $|e_{\vec{j}}\rangle_{\mathbb{H}_A^{\otimes n}}=|e_{j_1}\rangle_A\otimes\cdots\otimes|e_{j_n}\rangle_A$, respectively, where $\vec{j}=(j_1,\dots,j_n)$. The sequence $\vec{j}$ is called *strongly typical* if the numbers $N(j|\vec{j})$ of appearance of the symbol j in $\vec{j}$ satisfy the condition

$$\left|\frac{N(j|\vec{j})}{n}-\gamma_j\right|<\delta,\quad j=1,2,\dots,d,$$

and $N(j|\vec{j})=0$ if $\gamma_j=0$. Let us denote the collection of all strongly typical sequences as $B^{n,\delta}$, and let P^n be the probability distribution given by the eigenvalues $\gamma_{\vec{j}}$. Then by the law of large numbers, $P^n(B^{n,\delta})\to 1$ as $n\to+\infty$. It can be shown that the size of $B^{n,\delta}$ satisfies

$$2^{n(H(\rho)-\Delta_n(\delta))}<|B^{n,\delta}|<2^{n(H(\rho)+\Delta_n(\delta))}, \tag{13.63}$$

where $H(\rho)=-\sum_{i=1}^d\gamma_i\log\gamma_i$, and $\lim_{\delta\to 0}\lim_{n\to+\infty}\Delta_n(\delta)=0$. For the arbitrary function $f(j)$, $j=1,\dots,d$, and $\vec{j}=(j_1,\dots,j_n)\in B^{n,\delta}$, we have

$$\left|\frac{f(j_1)+\cdots+f(j_n)}{n}-\sum_{j=1}^d\gamma_jf(j)\right|<\delta\max_j f(j). \tag{13.64}$$

In particular, any strongly typical sequence is (entropy) typical: taking $f(j)=-\log\gamma_j$ gives

$$n(H(\rho)-\delta_1)<-\log\gamma_{\vec{j}}<n(H(\rho)+\delta_1), \tag{13.65}$$

where $\delta_1=\delta\max_{j:\gamma_j>0}(-\log\gamma_j)$. The converse is not true; -- not every typical sequence is strongly typical. The strongly typical projector is defined as the following spectral projector of $\rho^{\otimes n}$:

$$\mathbf{P}^{n,\delta} = \sum_{\vec{j}\in B^{n,\delta}} |e_{\vec{j}}\rangle\langle e_{\vec{j}}|.$$

We denote $d_{n,\delta} = \dim(\mathbf{P}^{n,\delta}) = |B^{n,\delta}|$ and $\bar{\rho}^{n,\delta} = \mathbf{P}^{n,\delta}$ and we are going to prove that

$$\lim_{\delta\to 0}\lim_{n\to+\infty}\frac{1}{n}H(\Psi^{\otimes n}(\bar{\rho}^{n,\delta})) = H(\Psi(\rho)) \tag{13.66}$$

for the arbitrary channel Ψ. We have

$$\begin{aligned} nH(\Psi(\rho)) - H(\Psi^{\otimes n}(\bar{\rho}^{n,\delta})) &= H\big((\Psi(\rho))^{\otimes n}\big) - H(\Psi^{\otimes n}(\bar{\rho}^{n,\delta})) \\ &= H(\Psi^{\otimes n}(\bar{\rho}^{n,\delta})\|\Psi^{\otimes n}(\rho^{\otimes n})) \\ &\quad + \operatorname{tr}\big[\log\big((\Psi(\rho))^{\otimes n}\big)\big(\Psi^{\otimes n}(\bar{\rho}^{n,\delta}) - (\Psi(\rho))^{\otimes n}\big)\big], \end{aligned} \tag{13.67}$$

where $H(\cdot\|\cdot)$ denotes a relative entropy. Strictly speaking, this formula is correct if the density operator $(\Psi(\rho))^{\otimes n}$ is nondegenerate, which we assume for a moment. Later we shall show how the argument can be modified to the general case. For the first term, we have the estimate by the fundamental property of monotonicity of the relative entropy

$$H(\Psi^{\otimes n}(\rho^{n,\delta})\|\Psi^{\otimes n}(\rho^{\otimes n})) \le H(\rho^{n,\delta}\|\rho^{\otimes n}),$$

with the right-hand side computed explicitly as

$$H(\rho^{n,\delta}\|\rho^{\otimes n}) = \sum_{\vec{j}\in B^{n,\delta}} \frac{1}{d_{n,\delta}}\log\left(\frac{1}{d_{n,\delta}}\gamma_{\vec{j}}\right) = -\log d_{n,\delta} - \sum_{\vec{j}\in B^{n,\delta}}\frac{1}{d_{n,\delta}}\log\gamma_{\vec{j}},$$

which is less than or equal to $n(\delta_1+\Delta_n(\delta))$ by (13.65), (13.63), giving a sufficient estimate. By using the identity,

$$\log\big((\Psi(\rho))^{\otimes n}\big) = \log\Psi(\rho)\otimes\mathbf{I}\otimes\cdots\otimes\mathbf{I} + \cdots + \mathbf{I}\otimes\cdots\otimes\mathbf{I}\otimes\log\Psi(\rho),$$

and introducing the operator $\mathfrak{F} = \Psi^*(\log\Psi(\rho))$ where $\Psi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ is the dual channel of Ψ, we can rewrite the second term of the right-hand side of (13.67) as

$$\begin{aligned} &n\operatorname{tr}\left[\frac{(\mathfrak{F}\otimes\mathbf{I}\otimes\cdots\otimes\mathbf{I}+\cdots+\mathbf{I}\otimes\cdots\otimes\mathbf{I}\otimes\mathfrak{F})}{n}(\bar{\rho}^{n,\delta}-\rho^{\otimes n})\right] \\ &\quad = \frac{n}{d_{n,\delta}}\sum_{\vec{j}\in B^{n,\delta}}\left(\frac{f(j_1)+\cdots+f(j_n)}{n} - \sum_{j=1}^{d}\gamma_j f(j)\right), \end{aligned}$$

where $f(j) = \langle e_j|\mathfrak{F}|e_j\rangle$, which is evaluated by $n\delta\max f$ via (13.64). This establishes (13.66) in the case of a nondegenerate $\Psi(\rho)$.

Coming back to the general case, let us denote $\mathbf{P}_\Psi$, the supporting projector of $\Psi(\rho)$. Then the supporting projector of $(\Psi(\rho))^{\otimes n}$ is $\mathbf{P}_\Psi^{\otimes n}$, and the support of $\Psi^{\otimes n}(\bar{\rho}^{n,\delta})$ is contained in the support of $(\Psi(\rho))^{\otimes n} = \Psi^{\otimes n}(\rho^{\otimes n})$, because the support of $\bar{\rho}^{n,\delta}$ is contained in the support of $\rho^{\otimes n}$. Thus, the second term in (13.67) should be understood as

$$\mathrm{tr}[\mathbf{P}_\Psi^{\otimes n}\log(\mathbf{P}_\Psi^{\otimes n}(\Psi(\rho))^{\otimes n}\mathbf{P}_\Psi^{\otimes n})\mathbf{P}_\Psi^{\otimes n}(\Psi^{\otimes n}(\bar{\rho}^{n,\delta}) - (\Psi(\rho))^{\otimes n})],$$

where now we have log of a nondegenerate operator in $\mathbf{P}_\Psi^{\otimes n}(\mathbb{H}_A^{\otimes n})$. We can then repeat the argument with $\mathfrak{F}$ defined as $\Psi^*(\mathbf{P}_\Psi(\log \mathbf{P}_\Psi\Psi(\rho)\mathbf{P}_\Psi)\mathbf{P}_\Psi)$. This fulfills the proof of (13.62), from which (13.59) follows. This proves the lemma. □

Lemma 13.4.3. *Assume that all conditions of Theorem* 13.4.1 *hold. Then*

$$C_{\mathrm{ea}}(\Phi) \le \max_{\omega_A} I_m(\omega_A;\Phi). \tag{13.68}$$

Proof. We now prove the inequality

$$C_{\mathrm{ea}}(\Phi) \le \max_{\omega_A} I_m(\omega_A;\Phi).$$

We first prove $C_{\mathrm{ea}}^{(1)}(\Phi) \le \max_{\omega_A} I_m(\omega_A;\Phi)$ as follows. Let us denote $\mathfrak{E}_A^\lambda$ the encodings used by A. Let ω_{AB} be the pure state initially shared by A and B, then the state after the encoding of the system AB and A are

$$\omega_{AB}^\lambda = (\mathfrak{E}_A^\lambda \otimes \mathfrak{I}_B)(\omega_{AB}) \quad \text{and} \quad \omega_A^\lambda = \mathfrak{E}_A^\lambda(\omega_A), \quad \text{respectively,} \tag{13.69}$$

where $\omega_A = \mathrm{tr}_B[\omega_{AB}]$. Note that the partial state of B does not change after the encoding, $\omega_B^\lambda = \omega_B$. We are going to prove that

$$H\Big(\sum_\lambda p_\lambda(\Phi\otimes\mathfrak{I}_B)(\omega_{AB}^\lambda)\Big) - \sum_\lambda p_\lambda H((\Phi\otimes\mathfrak{I}_B)(\omega_{AB}^\lambda)) \le I_m\Big(\sum_\lambda p_\lambda\omega_A^\lambda;\Phi\Big) \tag{13.70}$$

By the quantum coding theorem, the maximum of the left-hand side with respect to all possible p_λ, $\mathfrak{E}_A^\lambda$ and ω_{AB}^λ is just $C_{\mathrm{ea}}^{(1)}(\Phi)$. Therefore, $C_{\mathrm{ea}}^{(1)}(\Phi) \le \max_{\omega_A} I_m(\omega_A,\Phi)$.

By using subadditivity of quantum entropy, we can evaluate the first term on the left-hand side of (13.70) as

$$H\Big(\sum_\lambda p_\lambda\Phi(\omega_A^\lambda)\Big) + H(\omega_B) = H\Big(\Phi\Big(\sum_\lambda p_\lambda\omega_A^\lambda\Big)\Big) + \sum_\lambda p_\lambda H(\omega_B).$$

Here, the first term already gives the output entropy from $I_m(\sum_\lambda p_\lambda\omega_A^\lambda;\Phi)$. Let us proceed with evaluation of the remainder

$$\sum_\lambda p_\lambda(H(\omega_B) - H((\Phi\otimes\mathfrak{I}_B)(\omega_{AB}^\lambda)).$$

We first show that the term in the above-squared brackets does not exceed $H(\omega_A^\lambda) - H((\Phi \otimes \mathfrak{I}_{R^\lambda})(\omega_{AR^\lambda}^\lambda))$, where R^λ is the purifying (reference) system for ω_A^λ, and $\omega_{AR^\lambda}^\lambda$ is the purified state. To this end, consider the unitary extension of encoding $\mathfrak{E}_A^\lambda$ with the environment E^λ, which is initially in a pure state. From (13.69), we see that we can take $R^\lambda = BE_\lambda$ (after the unitary interaction which involves only AE_λ). Then, again denoting with primes, the states after the application of the channel Φ, we have

$$H(\omega_B) - H(\Phi \otimes \mathfrak{I}_B)(\omega_{AB})) = H(\omega_B) - H(\omega_{A'B}^\lambda) = -H(A|B)_{\omega_{A'B}^\lambda}. \tag{13.71}$$

Similarly,

$$\begin{aligned} &H(\omega_A) - H((\Phi \otimes \mathfrak{I}_{R^\lambda})(\omega_{AR^\lambda}^\lambda) \\ &\quad = H(\omega_{R^\lambda}^\lambda) - H(\omega_{A'R^\lambda}^\lambda) = -H(A'|R^\lambda)_{\omega_{A'B}^\lambda} = -H(A'|BE_\lambda)_{\omega_{A'B}^\lambda}, \end{aligned}$$

which is greater or equal than (13.71) by monotonicity of the conditional entropy.

We claim that the function $\omega_A \mapsto H(\omega_A) - H(\Phi \otimes \mathfrak{I}_R)(\omega_{AR}))$ is concave.

To prove this claim, we introduce an environment E for channel Φ and we have

$$\begin{aligned} H(\omega_A) - H((\Phi \otimes \mathfrak{I}_R)(\omega_{AR}) &= H(\omega_R) - H(\omega_{A'R}) \\ &= H(\omega_{A'E'}) - H(\omega_{E'}) = H(A'|E'). \end{aligned}$$

The conditional entropy $H(A'|E')$ is a concave function of $\omega_{A'E'}$. The map $\omega_A \mapsto \omega_{A'E'}$ is affine and, therefore, $H(A'|E')$ is a concave function of ω_A. This proves the claim that the function $\omega_A \mapsto H(\omega_A) - H(\Phi \otimes \mathfrak{I}_R)(\omega_{AR}))$ is concave.

Consequently,

$$\begin{aligned} &\sum_\lambda p_\lambda H(\omega_A^\lambda) - H((\Phi \otimes \mathfrak{I}_{R^\lambda})(\omega_{AR^\lambda}^\lambda)) \\ &\quad \le H\Big(\sum_\lambda p_\lambda \omega_A^\lambda\Big) - H((\Phi \otimes \mathfrak{I}_R)(\hat{\omega}_{AR})) \end{aligned}$$

where $\hat{\omega}_{AR}$ is the state purifying $\sum_\lambda \omega_A^\lambda$ with a reference system R.

Applying the same argument to the channel $\Phi^{\otimes n}$ gives

$$C_{\mathrm{ea}}^{(n)}(\Phi) \le \max_{\omega_A^{(n)}} I_m(\omega_A^{(n)}, \Phi^{\otimes n}). \tag{13.72}$$

Then from subadditivity of quantum mutual information (see Proposition 11.1.4), we have the remarkable additivity property

$$\max_{\omega_A^{(n)}} I_m(\omega_A^{(n)}; \Phi^{(n)}) = n \max_{\omega_A} I_m(\omega_A; \Phi).$$

Therefore, we obtain $C_{\mathrm{ea}}(\Phi) \le \max_{\omega_A} I_m(\omega_A; \Phi)$.

Lemmas (13.4.2) and (13.4.3) show that $C_{\mathrm{ea}}(\Phi) = \max_{\omega_A} I_m(\omega_A; \Phi)$. This proves Theorem 13.4.1. □

13.4.2 Constrained infinite-dimensional channels

In this subsection, the BSST theorem for finite-dimensional input and output Hilbert spaces presented in the previous subsection is extended to the infinite case with some linear constraints described below.

When applying the protocol of entanglement-assisted communication to infinite-dimensional channels, one has to impose certain constraints on the input states. Otherwise, it can be shown that there exists the unconstrained memoryless channel Φ with $C_{\mathrm{ea}}(\Phi) = +\infty$.

A typical physically motivated constraint is the bounded energy of states used for encoding. This constraint is determined by the linear inequality

$$\mathrm{tr}[\rho \mathbf{H}] \le E, \quad E > 0, \tag{13.73}$$

where $\mathbf{H}$ is a positive self-adjoint operator—a Hamiltonian of the input quantum system $\mathbb{H}_A$, which can be assumed to be a $\mathfrak{H}$-operator. In this case, we again denote the compact subset $\mathcal{K}_{\mathbf{H}}(E)$ of $\mathcal{S}(\mathbb{H}_A)$ (see (3.3) for compactness of $\mathcal{K}_{\mathbf{H}}(E)$ in $\mathbb{H}_A$) as

$$\mathcal{K}_{\mathbf{H}}(E) = \{\rho \in \mathcal{S}(\mathbb{H}_A) \mid \mathrm{tr}[\rho \mathbf{H}] \le E\}.$$

For the $\mathfrak{H}$-operator $\mathbf{H}$ on the infinite-dimensional space $\mathbb{H}_A$ and any state $\rho \in \mathcal{S}(\mathbb{H}_A)$, the energy $\mathrm{tr}[\rho \mathbf{H}]$ (finite or infinite) is defined as $\sup_n \mathrm{tr}[\rho \mathbf{P}_n \mathbf{H} \mathbf{P}_n]$, where $\mathbf{P}_n$ is the finite-dimensional spectral projector of $\mathbf{H}$ corresponding to the interval $[0, n]$.

We impose the following linear constraint onto the input states $\rho^{(n)}$ of the channel $\Phi^{\otimes n}$

$$\mathrm{tr}[\rho^{(n)} \mathbf{H}^{(n)}] \le nE, \tag{13.74}$$

where

$$\mathbf{H}^{(n)} = \mathbf{H} \otimes \mathbf{I}_A \otimes \cdots \otimes \mathbf{I}_A + \cdots + \mathbf{I}_A \otimes \cdots \otimes \mathbf{I}_A \otimes \mathbf{H}. \tag{13.75}$$

In this case, the constrained Holevo χ-capacity $C_\chi(\Phi; \mathbf{H}, E)$ is defined by

$$C_\chi(\Phi; \mathbf{H}, E) = \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} C_\chi(\rho, \Phi), \tag{13.76}$$

where

$$C_\chi(\rho, \Phi) = \sup_{\sum_i p_i \rho_i = \rho} \chi_\Phi(\{p_i, \rho_i\}), \tag{13.77}$$

is the constrained χ-capacity of the channel Φ at the state ρ (the supremum is over all ensembles with the average state $\sum_i p_i\rho_i = \rho$). If $H(\Phi(\rho)) < +\infty$, then

$$C_\chi(\rho, \Phi) = H(\Phi(\rho)) - \hat{H}_\Phi(\rho), \tag{13.78}$$

where $\hat{H}_\Phi(\rho) = \inf_{\sum_i p_i\rho_i=\rho} H(\Phi(\rho))$ is the σ-convex hull of the function $\rho \mapsto H(\Phi(\rho))$. Due to the concavity of the von Neumann entropy $H(\cdot)$, the infimum can be taken over ensembles of pure states because any mixed quantum state can be expressed as a convex combination of pure states. By the Holevo–Schumacher–Westmoreland (HSW) theorem adapted to constrained channels (see (13.47)), the classical capacity of the channel Φ with constraint (16.92) is given by the following regularized expression:

$$C(\Phi; \mathbf{H}, E) = \lim_{n\to+\infty} \frac{1}{n} C_\chi(\Phi^{\otimes n}; \mathbf{H}^{(n)}, nE),$$

where $\mathbf{H}^{(n)}$ is defined by (13.75).

Constraint (13.74) is equivalent to a similar constraint on input states of the channel $\Phi^{\otimes n} \otimes \mathfrak{I}_B^{\otimes n}$ with the constraint operator $\mathbf{H}_{AB}^{(n)} = \mathbf{H}^{(n)} \otimes \mathbf{I}_B^{(n)}$ on the composite Hilbert space $\mathbb{H}_{AB}^{\otimes n}$, where $\mathbf{I}_B^{(n)}$ is the identity operator on $\mathbb{H}_B^{\otimes n}$. Denote by $\mathcal{P}_{AB}^{(n)}$ the collection of ensembles $\pi^{(n)} = \{p_\lambda^{(n)}, \omega_\lambda^{(n)}\}$, where $\omega_\lambda^{(n)}$ are states of the form (13.73) satisfying

$$\sum_{\lambda\in\Lambda} p_\lambda^{(n)} \operatorname{tr}[\omega_\lambda^{(n)} \mathbf{H}_{AB}^{(n)}] \le nE.$$

The classical capacity of the above protocol is called the *entanglement-assisted classical capacity* of the channel Φ under constraint (13.73) and is denoted by $C_{\mathrm{ea}}(\Phi; \mathbf{H}, E)$. In the following, we obtain

$$C_{\mathrm{ea}}(\Phi; \mathbf{H}, E) = \lim_{n\to+\infty} \frac{1}{n} C_{\mathrm{ea}}^{(n)}(\Phi; \mathbf{H}, E), \tag{13.79}$$

where

$$C_{\mathrm{ea}}^{(n)}(\Phi; \mathbf{H}, E) = \sup_{\pi^{(n)}\in\mathcal{P}_{AB}^{(n)}} \chi_{\Phi^{\otimes n}\otimes\mathfrak{I}_B^{\otimes n}}(\{p_\lambda^{(n)}, \rho_\lambda^{(n)}\}). \tag{13.80}$$

To establish the main result for the entanglement-assisted classical capacity for the memoryless channel, we recall the definition and property of mutual information $I_m(\rho, \Phi)$ below.

For finite dimensions, it is defined for arbitrary state $\rho \in \mathcal{S}(\mathbb{H}_A)$,

$$I_m(\rho, \Phi) = H(\rho) + H(\Phi(\rho)) - H((\Phi \otimes \mathfrak{I}_R)(\hat{\rho})), \tag{13.81}$$

where $\mathbb{H}_R$ is a Hilbert space isomorphic to $\mathbb{H}_A$ and $\hat{\rho}$ is the purification of ρ in the space $\mathbb{H}_A \otimes \mathbb{H}_R$ so that $\operatorname{tr}_R[\hat{\rho}] = \rho$. By the complementary channel $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$, the quantum mutual information $I_m(\rho, \Phi)$ can also be written as

$$I_m(\rho,\Phi) = H(\rho) + H(\Phi(\rho)) - H(\hat{\Phi}(\rho)). \tag{13.82}$$

In the infinite dimension expressions (13.81), (13.82) may contain uncertainty "$\infty-\infty$," and to avoid this problem they should be modified as

$$I_m(\rho,\Phi) = H((\Phi\otimes\mathfrak{I}_R)(\hat{\rho})\|(\Phi\otimes\mathfrak{I}_R)(\rho\otimes\rho)), \tag{13.83}$$

where $\rho = \mathrm{tr}_A[\hat{\rho}]$ is the state in $\mathcal{S}(\mathbb{H}_R)$ with the same nonzero spectrum as ρ. Analytical properties of the function $(\rho,\Phi)\mapsto I_m(\rho,\Phi)$ were studied in Theorem 11.1.9 in the infinite-dimensional case.

We also review property of conditional entropy $H(\cdot|\cdot)$ below.

In finite dimensions, the conditional entropy of a state ρ of a composite system AB is defined as

$$H(A|B)_\rho := H(\rho) - H(\rho_B), \tag{13.84}$$

where $\rho_B = \mathrm{tr}_A[\rho]$. The conditional entropy is finite, but in contrast to the classical case it may be negative.

The conditional entropy of a state ρ of an infinite-dimensional composite system AB is defined as follows (see Definition 8.3.2):

$$H(A|B)_\rho := H(\rho_A) - H(\rho\|\rho_A\otimes\rho_B), \tag{13.85}$$

provided $H(\rho_A) < +\infty$. It is easy to see that the right-hand sides of (13.84) and (13.85) coincide if $H(\rho) < +\infty$ (finiteness of any two values from the triple $(H(\rho_A), H(\rho_B), H(\rho))$ implies finiteness of the third one). It is proved in Proposition 8.3.6 that the above-defined conditional entropy is a concave function on the convex set of all states ρ of the system AB such that $H(\rho_A) < +\infty$, possessing the following properties:

$$H(A|B)_{\rho_{AB}} \geq H(A|BC)_\rho \tag{13.86}$$

for any state ρ of ABC (monotonicity), and

$$H(A|B)_{\rho_{AB}} = -H(A|C)_{\rho_{AC}} \tag{13.87}$$

for any pure state ρ of ABC, where it is assumed that $H(\rho_A) < +\infty$.

With the above facts in mind, we have the following result.

Lemma 13.4.4. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be a quantum channel and σ an arbitrary state in $\mathcal{S}(\mathbb{H}_B)$. Then for an arbitrary ensemble $\{p_i,\omega_i\}$ of states in $\mathcal{S}(\mathbb{H}_{AB})$, where $\mathbb{H}_{AB} = \mathbb{H}_A\otimes\mathbb{H}_B$, such that $(\omega_i)_B := \mathrm{tr}_A[\omega_i] = \sigma \in \mathcal{S}(\mathbb{H}_B)$ for all i, the inequality*

$$\chi_{\Phi\otimes\mathfrak{I}_B}(\{p_i,\omega_i\}) \leq I_m(\omega_A,\Phi), \tag{13.88}$$

holds, where $\omega = \sum_i p_i\omega_i$ is the average state of the ensemble $\{p_i,\omega_i\}$ and $\omega_A = \mathrm{tr}_B[\omega]$.

Proof. Let $\{p_i, \omega_i\}$ be an ensemble of states in $\mathcal{S}(\mathbb{H}_{AB})$ with an average state ω such that $(\omega_i)_B = \sigma \in \mathcal{S}(\mathbb{H}_B)$ for all i. We have to show that

$$\sum_i p_i H\big((\Phi \otimes \mathfrak{I}_B)(\omega_i) \| (\Phi \otimes \mathfrak{I}_B)(\omega)\big) \leq I_m(\omega_A, \Phi). \tag{13.89}$$

First, let us prove inequality (13.88) assuming that Φ is an *FF* channel, i. e., $\dim(\mathbb{H}_A) < +\infty$ and $\dim(\mathbb{H}_{A'}) < +\infty$. In this case, the left-hand side of this inequality can be rewritten as

$$\begin{aligned}\chi_{\Phi\otimes\mathfrak{I}_B}(\{p_i, \omega_i\}) &= \sum_i p_i H\big((\Phi \otimes \mathfrak{I}_B)(\omega_i) \| (\Phi \otimes \mathfrak{I}_B)(\omega)\big) \\ &= H\big((\Phi \otimes \mathfrak{I}_B)(\omega)\big) - \sum_i p_i H\big((\Phi \otimes \mathfrak{I}_B)(\omega_i)\big) \\ &\leq H\big(\Phi(\mathrm{tr}_B[\omega])\big) + H(\mathfrak{I}_B)(\mathrm{tr}_A[\omega]) - \sum_i p_i H\big((\Phi \otimes \mathfrak{I}_B)(\omega_i)\big) \\ &= H\big(\Phi(\omega_A)\big) + \sum_i p_i\big(H(\sigma) - H\big((\Phi \otimes \mathfrak{I}_B)(\omega_i)\big)\big),\end{aligned}$$

where the inequality in the above is due to subadditivity of the von Neumann entropy. Note that $H((\Phi \otimes \mathfrak{I}_B)(\omega_i)) - H(\sigma)$ is the conditional entropy $H(A'|B)$ at the state $(\Phi \otimes \mathfrak{I}_B)(\omega_i)$. Let $\hat{\omega}_i$ be a pure state in ABR_i such that $\mathrm{tr}_{R_i}[\hat{\omega}_i] = \omega_i$. By monotonicity of the conditional entropy (property (13.86)), we have

$$H\big((\Phi \otimes \mathfrak{I}_B)(\omega_i)\big) - H(\sigma) = H(A'|B)_{(\Phi\otimes\mathfrak{I}_B)(\omega_i)} \geq H(A'|BR_i)_{(\Phi\otimes\mathfrak{I}_{BR_i})(\hat{\omega}_i)}, \tag{13.90}$$

where $H(A'|BR_i)$ is defined by (13.85) (the system R_i is infinite-dimensional, but the system A' is finite-dimensional by the assumption). Since $\hat{\omega}_i$ is a purification of the state $\rho_i := (\omega_1)_A$, i. e., $(\hat{\omega}_1)_A = \rho_i$, property (13.87) of the conditional entropy implies

$$\begin{aligned}H(A'|BR_i)_{\Phi\otimes\mathfrak{I}_{BR_i}(\hat{\omega}_i)} &= H(A'|BR_i)_{\mathrm{tr}_E[(V\otimes\mathfrak{I}_{BR_i})\hat{\omega}_i(\mathbf{V}^*\otimes\mathfrak{I}_{BR_i})]} \\ &= -H(A'|E)_{\mathrm{tr}_{BR_i}[(\mathbf{V}\otimes\mathfrak{I}_{BR_i})\hat{\omega}_i(\mathbf{V}^*\otimes\mathfrak{I}_{BR_i})]} = -H(A'|E)_{\mathbf{V}\rho_i\mathbf{V}^*},\end{aligned} \tag{13.91}$$

where E is an environment system for the channel Φ and $\mathbf{V}$ is the Stinespring isometry (i. e., $\Phi(\rho) = \mathrm{tr}_E[\mathbf{V}\rho\mathbf{V}^*]$). By using concavity of the conditional entropy (defined by (13.85)) and property (13.87), we obtain

$$\sum_i p_i H(A'|E)_{\mathbf{V}\rho_i\mathbf{V}^*} \leq H(A'|E)_{\mathbf{V}\rho\mathbf{V}^*} = -H(A'|R)_{\mathrm{tr}_E[(\mathbf{V}\otimes\mathbf{I}_R)\hat{\rho}(\mathbf{V}^*\mathbf{I}_R])},$$

where R is a reference system for the state ρ and $\hat{\rho}$ is a pure state in AR such that $\hat{\rho}_A$, and (13.90) and (13.91) imply

$$\sum_i p_i H\big((\Phi\otimes\mathfrak{I}_B)(\omega_i)\|(\Phi\otimes\mathfrak{I}_B)(\omega)\big) \le H\big(\Phi(\rho)\big) - H(A'|R)_{(\Phi\otimes\mathfrak{I}_R)(\hat\rho)}$$
$$= H\big((\Phi\otimes\mathfrak{I}_R)(\hat\rho)\|\Phi(\rho)\otimes\hat\rho_R\big) = I_m(\rho,\Phi),$$

where Definitions 13.83 and 13.85 were used. Thus, inequality (13.88) is proved under the assumption that $\dim(\mathbb{H}_{A'}) < +\infty$, $\dim(\mathbb{H}_B) < +\infty$.

Its proof in the general case can be obtained using approximation techniques as follows. Let $\{p_i,\omega_i\}$ be an ensemble such that $(\omega_i)_B := \mathrm{tr}_A[\omega_i] = \sigma \in \mathcal{S}(\mathbb{H}_B)$, and let $\mathbf{Q}_n$ be the spectral projector of the state σ corresponding to its n maximal eigenvalues. Let $\lambda_n = \mathrm{tr}[\mathbf{Q}_n\sigma]$ and $\mathbf{C}_n = \mathbf{I}_A\otimes\mathbf{Q}_n$. For $n\in\mathbb{N}$, consider the ensemble $\{p_i,\omega_i^n\}$ with the average state ω^n, where

$$\omega_i^n = \lambda_n^{-1}\mathbf{C}_n\omega_i\mathbf{C}_n, \quad \omega^n = \lambda_n^{-1}\mathbf{C}_n\omega\mathbf{C}_n.$$

Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be a sequence of finite-rank projectors in the space $\mathbb{H}_{A'}$ strongly converging to the identity operator $\mathbf{I}_{A'}$, and let τ be a pure state in $\mathcal{S}(\mathbb{H}_{A'})$. Consider the sequence of channels $\Phi_n = \Pi_n\circ\Phi$, where

$$\Pi_n(\Phi)(\rho) = \mathbf{P}_n\rho\mathbf{P}_n + \tau'\,\mathrm{tr}\big[(\mathbf{I}_{A'}-\mathbf{P}_n)\rho\big], \quad \rho\in\mathcal{S}(\mathbb{H}_{A'}).$$

Since $(\omega_i^n)_B := \mathrm{tr}_A[\omega^n] = \lambda_n^{-1}\mathbf{Q}_n\sigma$ for all i, the first part of the proof implies

$$\sum_i p_i H\big((\Phi_n-\mathfrak{I}_B)(\omega_i^n)\|(\Phi_n-\mathfrak{I}_B)(\omega^n)\big) \le I_m(\omega_A^n,\Phi_n).$$

Since $\lambda_n\omega_A^n \le \omega_A$, we therefore have $\lim_{n\to+\infty} I_m(\omega_A^n,\Phi_n) = I_m(\omega_A,\Phi)$. Hence, the above inequality implies inequality (13.88) by lower semicontinuity of the relative entropy. This proves the lemma. □

The following result is due originally to Holevo and Shirokov [83].

Theorem 13.4.5. *Let* $\Phi:\mathcal{S}(\mathbb{H}_A)\to\mathcal{S}(\mathbb{H}_{A'})$ *be a quantum channel, and let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on* $\mathbb{H}_A$. *The entanglement-assisted classical capacity (finite or infinite) of the channel* Φ *with constraint* (13.73) *is given by the expression*

$$C_{\mathrm{ea}}(\Phi;\mathbf{H},E) = \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} I_m(\rho;\Phi). \tag{13.92}$$

We prove the above theorem via the following three lemmas.

Lemma 13.4.6. *Assume that* $\dim(\mathbb{H}_{A'}) < +\infty$. *Let* $\Phi:\mathcal{S}(\mathbb{H}_A)\to\mathcal{S}(\mathbb{H}_{A'})$ *be a quantum channel, and let* $\mathbf{H}$ *be a positive self-adjoint linear operator* $\mathbb{H}_A$. *Then the entanglement-assisted classical capacity (finite or infinite) of the channel* Φ *with constraint* (13.73) *satisfies the following inequality:*

$$C_{\mathrm{ea}}(\Phi;\mathbf{H},E) \ge \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} I_m(\rho;\Phi). \tag{13.93}$$

Proof. Assuming that $\dim(\mathbb{H}_{A'}) < +\infty$, we observe the following:
1. Finite dimensionality of the system $\mathbb{H}_{A'}$ implies finiteness of the output entropy of the channel Φ on the whole space of input states $\mathcal{S}(\mathbb{H}_A)$. That is, $H(\Phi(\rho)) < +\infty$ for all $\rho \in \mathcal{S}(\mathbb{H}_A)$;
2. Finiteness of $\mathrm{tr}[\rho\mathbf{H}]$ implies that all the eigenvectors of the state ρ belong to the domain of the operator $\sqrt{\mathbf{H}}$.
3. Finite dimensionality of the system $\mathbb{H}_{A'}$ shows that for any finite-rank state ρ the restriction of the channel $\Phi^{\otimes n}$ to the support of the state $\rho^{\otimes n}$ acts as a finite-dimensional channel for each n;
4. If there are no states ρ satisfying the inequality $\mathrm{tr}[\rho\mathbf{H}] < E$ but there exists an infinite-rank state ρ_0 such that $\mathrm{tr}[\rho_0\mathbf{H}] = E$, then there is a sequence $(\rho_n)_{n=1}^{+\infty}$ of finite-rank states converging to ρ_0 such that $\mathrm{tr}[\rho_n\mathbf{H}] = E$ for which

$$\liminf_{n\to+\infty} I_m(\rho_n,\Phi) \geq I_m(\rho_0,\Phi)$$

by lower semicontinuity of the quantum mutual information.

With the observations above, the rest of the proof of this lemma follows similar to that of Lemma 13.4.2. This proves the lemma. □

Lemma 13.4.7. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be a quantum channel, and let $\mathbf{H}$ be an $\mathfrak{H}$-operator on $\mathbb{H}_A$. The entanglement-assisted classical capacity (finite or infinite) of the channel Φ with constraint* (13.73) *is given by the expression*

$$C_{\mathrm{ea}}(\Phi;\mathbf{H},E) \geq \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} I_m(\rho;\Phi). \tag{13.94}$$

Proof. Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be an arbitrary channel. We prove that $C_{\mathrm{ea}}(\Phi;\mathbf{H},E) \geq \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} I_m(\rho;\Phi)$ as follows.

Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be sequence of finite-dimensional projectors on $\mathbb{H}_{A'}$ strongly converging to the unit operator $\mathbf{I}_{A'}$ on $\mathbb{H}_{A'}$. The channel Φ is approximated in the strong convergence topology by the sequence $(\Pi_n\circ\Phi)_{n=1}^{+\infty}$ with finite-dimensional output, where $\Pi_n(\sigma) = \mathbf{P}_n\sigma\mathbf{P}_n + (\mathrm{tr}[\sigma(\mathbf{I}_{A'} - \mathbf{P}_n)])\tau$ for all $\sigma \in \mathcal{S}(\mathbb{H}_{A'})$ and τ is a given state in $\mathbb{H}_{A'}$. Since the inequality "$\geq$" in (13.93) is proved for a channel with finite-dimensional output (see Lemma 13.4.6), the chain rule for the entanglement-assisted capacity implies

$$C_{\mathrm{ea}}(\Phi;\mathbf{H},E) \geq C_{\mathrm{ea}}(\Pi_n\circ\Phi;\mathbf{H},E) \geq I_m(\rho;\Pi_n\circ\Phi).$$

Lower semicontinuity of the function $\Phi \mapsto I_m(\rho;\Phi)$ in the strong convergence topology and the chain rule for quantum mutual information (see Proposition 11.2.3) imply

$$\lim_{n\to+\infty} I_m(\rho;\Pi_n\circ\Phi) = I_m(\rho;\Phi) \leq +\infty, \quad \forall\rho.$$

Hence, the inequality "≥" in (13.93) for the channel Φ follows from the above inequality. This proves the lemma. □

Lemma 13.4.8. *Let* $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ *be a quantum channel, and let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on* $\mathbb{H}_A$*. The entanglement-assisted classical capacity (finite or infinite) of the channel* Φ *with constraint* (13.73) *is given by the expression*

$$C_{\mathrm{ea}}(\Phi;\mathbf{H},E) \leq \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} I_m(\rho;\Phi). \tag{13.95}$$

Proof. We prove the inequality (13.95). For the ensemble $\{p_\lambda,\omega_\lambda\}$ of encoded states in $\mathcal{S}(\mathbb{H}_{AA'})$. Let $(\omega_\lambda)_A := \mathrm{tr}_{A'}[\omega_\lambda] = \sigma \in \mathcal{S}(\mathbb{H}_A)$, where $\mathrm{tr}_{A'}[\cdots]$ denotes the partial trace of $[\cdots]$ taken over $\mathbb{H}_{A'}$. By Lemma 13.4.4, we have the following inequality:

$$\chi_{\Phi^{\otimes n}\otimes\mathfrak{I}_B^{\otimes n}}(\{p_\lambda^{(n)},\omega_\lambda^{(n)}\}) \leq I_m\Bigl(\sum_\lambda p_\lambda^{(n)}(\omega_\lambda^{(n)})_A,\Phi^{\otimes n}\Bigr).$$

From (13.79), we have

$$C_{\mathrm{ea}}(\Phi,\mathbf{H},E) \leq \lim_{n\to+\infty}\frac{1}{n}\sup_{\pi^{(n)}\in\mathcal{P}_{AB}^{(n)}} I_m\Bigl(\sum_\lambda p_\lambda^{(n)}(\omega_\lambda^{(n)})_A,\Phi^{\otimes n}\Bigr).$$

Now

$$\begin{aligned}&\sup_{\pi^{(n)}\in\mathcal{P}_{AB}^{(n)}} I_m\Bigl(\sum_\lambda p_\lambda^{(n)}(\omega_\lambda^{(n)})_A,\Phi^{\otimes n}\Bigr)\\ &\quad\leq \sup_{\rho^{(n)}:\mathrm{tr}[\rho^{(n)}\mathbf{H}^{(n)}]\leq nE} I_m(\rho^{(n)},\Phi^{\otimes n}) \equiv \bar{I}_m^{(n)}(\Phi).\end{aligned}$$

We claim that the sequence $(\bar{I}_m^{(n)}(\Phi))_{n=1}^{+\infty}$ is additive. To show that, it suffices to prove that

$$\bar{I}_m^{(n)}(\Phi) \leq n\bar{I}_m^{(1)}(\Phi). \tag{13.96}$$

By the subadditivity of quantum mutual information,

$$I_m(\rho^{(n)},\Phi^{\otimes n}) \leq \sum_j I_m(\rho_j^{(n)},\Phi),$$

where $\rho_j^{(n)}$ are partial states, and by concavity,

$$\sum_j I_m(\rho_j^{(n)},\Phi) \leq nI_m\Bigl(\sum_j \rho_j^{(n)},\Phi\Bigr).$$

The inequality $\mathrm{tr}[\rho^{(n)}\mathbf{H}^{(n)}] \leq nE$ is equivalent to $\mathrm{tr}[(\frac{1}{n}\sum_{j=1}^n \rho_j^{(n)})\mathbf{H}] \leq E$; hence, (13.96) holds. Thus,

$$C_{\mathrm{ea}}(\Phi, \mathbf{H}, E) \leq \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} I_m(\rho; \Phi).$$

This proves the lemma. □

Lemmas 13.4.6, 13.4.7 and 13.4.8 together prove that

$$C_{\mathrm{ea}}(\Phi, \mathbf{H}, E) = \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} I_m(\rho; \Phi).$$

Therefore, Theorem 13.4.5 follows.

The following corollary for unconstrained channels follows immediately from the above theorem.

Corollary 13.4.9. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a memoryless quantum channel. Then the entanglement-assisted classical capacity (finite or infinite) of the unconstrained channel Φ is given by the expression*

$$C_{\mathrm{ea}}(\Phi) = \sup_{\rho \in \mathcal{S}(\mathbb{H}_A)} I_m(\rho; \Phi). \tag{13.97}$$

13.4.3 Continuity of $C_{\mathrm{ea}}(\cdot) : \mathfrak{QC}(A, B) \to [0, +\infty]$

Since a physical channel is always determined with some finite accuracy, it is necessary to explore the question of continuity of its information capacity with respect to small perturbations of a channel. Mathematically, this means that we have to study continuity of the capacity as a function of a channel assuming that the set of all channels is equipped with some appropriate topology. In this subsection, we consider continuity properties of the entanglement-assisted classical capacity with respect to the strong convergence topology on the set of all channels. Recall that strong convergence of a sequence of channels $(\Phi_n)_{n=1}^{+\infty}$, $\Phi_n : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ is said to converge strongly to a channel $\Phi_0 : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ if $\lim_{n\to+\infty} \Phi_n(\rho) = \Phi_0(\rho)$ for all states $\rho \in \mathcal{S}(\mathbb{H}_A)$.

Proposition 11.1.7 and lower semicontinuity of quantum mutual information $I_m(\cdot) : \mathfrak{QC}(A, B) \to [0, +\infty]$ as a function of a channel in the strong convergence topology imply that $\Phi \mapsto C_{\mathrm{ea}}(\Phi; \mathbf{H}, E)$ is a lower semicontinuous function in this topology on the set of all quantum channels; i. e.,

$$\liminf_{n\to+\infty} C_{\mathrm{ea}}(\Phi_n; \mathbf{H}, E) \geq C_{\mathrm{ea}}(\Phi_0; \mathbf{H}, E)$$

for any sequence $(\Phi_n)_{n=1}^{+\infty}$ of channels strongly converging to the channel Φ_0, i. e., $\lim_{n\to+\infty} \Phi_n(\rho) = \Phi_0(\rho)$ for all $\rho \in \mathcal{S}(\mathbb{H}_A)$.

The following proposition, due originally to Holevo and Shirokov [83], gives sufficient conditions for the continuity.

Proposition 13.4.10. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator on* $\mathbb{H}_A$ *such that* $\mathrm{tr}[\exp(-\lambda\mathbf{H})] < +\infty$ *for all* $\lambda > 0$*, and let* $(\Phi_n)_{n=1}^{+\infty}$ *be a sequence of quantum channels from A to A′ that converges strongly to a quantum channel* Φ_0*. Then the following relation:*

$$\lim_{n\to+\infty} C_{\mathrm{ea}}(\Phi_n;\mathbf{H},E) = C_{\mathrm{ea}}(\Phi_0;\mathbf{H},E) < +\infty \tag{13.98}$$

holds if one of the following conditions is met:
1. $\lim_{n\to+\infty} H(\Phi_n(\rho_n)) = H(\Phi_0(\rho_0))$ *for any sequence of states* $(\rho_n)_{n=1}^{+\infty}$ *in* $\mathcal{S}(\mathbb{H}_A)$ *that converges to* ρ_0 *under* $\|\cdot\|_1$*-norm such that* $\mathrm{tr}[\rho_n\mathbf{H}] < E$ *for* $n = 0, 1, 2, \ldots$.
2. *There exists a sequence of channels* $(\hat{\Phi}_n)_{n=1}^{+\infty}$ *from A to E that converges strongly to a channel* $\hat{\Phi}_0$ *such that* $(\Phi_n, \hat{\Phi}_n)$ *forms a complementary pair of channels for each* $n = 0, 1, 2, \ldots$.

Proof. We first note that the set

$$\mathcal{K}_{\mathbf{H}}(E) = \{\rho \in \mathcal{S}(\mathbb{H}_A) \mid \mathrm{tr}[\rho\mathbf{H}] < E\}$$

is a compact subset of $\mathcal{S}(\mathbb{H}_A)$ by Theorem 3.2.5 and the function $\rho \mapsto H(\rho)$ is continuous on this set, since $\mathrm{tr}[\exp(-\lambda\mathbf{H})] < +\infty$ for all $\lambda > 0$ by Proposition 7.3.7. By Proposition 11.1.7, the function $\rho \mapsto I_m(\rho;\Phi_n)$ is continuous on the compact set $\mathcal{K}_{\mathbf{H}}(E)$ for each $n \in \mathbb{N}$, and hence,

$$C_{\mathrm{ea}}(\Phi_n;\mathbf{H},E) = \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} I_m(\rho;\Phi_n) = \lim_{n\to+\infty} I_m(\rho_n;\Phi_n) < +\infty$$

for a particular sequence of states $(\rho_n)_{n=1}^{+\infty}$ in $\mathcal{K}_{\mathbf{H}}(E)$. Assume that there exists $\Phi_0 \in \mathfrak{QC}(A, A')$ such that the sequence $(\Phi_n)_{n=1}^{+\infty}$ converges strongly to Φ_0 and yet

$$\lim_{n\to+\infty} C_{\mathrm{ea}}(\Phi_n;\mathbf{H},E) > C_{\mathrm{ea}}(\Phi_0;\mathbf{H},E) \tag{13.99}$$

for contradiction purpose. Since $C_{\mathrm{ea}}(\cdot;\mathbf{H},E) : \mathfrak{QC}(A,B) \to [0,+\infty]$ is lower semicontinuous under the strong convergence topology, we have

$$\liminf_{n\to+\infty} C_{\mathrm{ea}}(\Phi_n;\mathbf{H},E) \geq C_{\mathrm{ea}}(\Phi_0;\mathbf{H},E).$$

Therefore, to show that

$$\lim_{n\to+\infty} C_{\mathrm{ea}}(\Phi_n;\mathbf{H},E) = C_{\mathrm{ea}}(\Phi_0;\mathbf{H},E),$$

it suffices to show that (13.99) leads to a contradiction.

Since $\mathcal{K}_{\mathbf{H}}(E)$ is a compact set, we may assume (by passing to a subsequence) that the sequence $(\rho_n)_{n=1}^{+\infty}$ converges to a particular state $\rho_0 \in \mathcal{K}_{\mathbf{H}}(E)$. Hence, to obtain a contradiction to (13.99), it suffices to prove that

$$\lim_{n\to+\infty} I_m(\rho_n;\Phi_n) = I_m(\rho_0;\Phi_0). \tag{13.100}$$

Conditions 1 and 2 of Proposition 13.4.10 provide two different ways to prove (13.99). If condition 1 holds, then

$$I_m(\rho_n;\Phi_n) = H(\rho_n) + H(\Phi_n(\rho_n)) - H((\Phi_n \otimes \mathfrak{I}_R)(|\phi_n\rangle_{AR}\langle\phi_n|)),$$

where $|\phi_n\rangle_{AR}$ is any purification for the state ρ_n, $n = 0, 1, 2, \ldots$ with a reference system R. By lower semicontinuity of the function $(\Phi,\rho) \mapsto I_m(\rho;\Phi)$, continuity of the entropy on the set $\mathcal{K}_{\mathbf{H}}(E)$, and condition 1, to prove (13.99) it suffices to show that

$$\liminf_{n\to+\infty} H((\Phi_n \otimes \mathfrak{I}_R)(|\phi_n\rangle_{AR}\langle\phi_n|)) \geq H((\Phi_0 \otimes \mathfrak{I}_R)(|\phi_0\rangle_{AR}\langle\phi_0|)).$$

This relation follows from lower semicontinuity of the relative entropy, since strong convergence of the sequence $(\Phi_n)_{n=1}^{+\infty}$ to the channel Φ_0 implies strong convergence of the sequence $(\Phi_n \otimes \mathfrak{I}_R)_{n=1}^{+\infty}$ to the channel $\Phi_0 \otimes \mathfrak{I}_R$, and we can choose a sequence of purifications of $|\phi_n\rangle_A$, $(|\phi_n\rangle_{AR})_{n=1}^{+\infty}$, that converges to the purification of $|\phi_0\rangle_A$, $|\phi_0\rangle_{AR}$. If condition 2 holds, then (13.99) directly follows from the continuity of mutual and coherent information (see Proposition 11.1.7). This proves the proposition. □

Remark 13.1.

1. While Gaussian quantum channels is not a subject of discussion in this book, we mention here that Condition 1 of Proposition 13.4.10 holds for any converging sequence of Gaussian channels provided that $\mathbf{H}$ is an oscillator Hamiltonian of a Bosonic system. Gaussian quantum channels are real examples of infinite-dimensional channels and is a currently a big research area that will have a huge impact on modern quantum communication. The reader can consult the works by Shirokov [148], De Palma et al. [36] and references contained therein for detailed discussion of Gaussian quantum channels.
2. Condition 2 of Proposition 13.4.10 holds for the sequence of the channels

   $$\Phi_n(\rho) = \sum_{i=1}^{+\infty} \mathbf{V}_i^n \rho \mathbf{V}_i^{n*},$$

   where $(\mathbf{V}_i^n)_{n=1}^{+\infty}$ is a sequence of operators from $\mathbb{H}_A$ to $\mathbb{H}_{A'}$ strongly converges to the operator $\mathbf{V}_i^0$ for each i such that $\sum_{i=1}^{+\infty}(\mathbf{V}_i^n)^*\mathbf{V}_i^n = \mathbf{I}_A$ for all n. Indeed,

   $$\hat{\Phi}_n(\rho) = \sum_{i,j=1}^{+\infty} \operatorname{tr}[(\mathbf{V}_i^n \rho \mathbf{V}_i^{n*})(|i\rangle_E\langle j|)],$$

   where $(|i\rangle_E)_{i=1}^{+\infty}$ is an orthonormal basis in $\mathbb{H}_E$, and it is easy to see the sequences $(\Phi_n)_{n=1}^{+\infty}$ and $(\hat{\Phi})_{n=1}^{+\infty}$ strongly converges, respectively, to Φ_0 and $\hat{\Phi}_0$ (Φ_0 and $\hat{\Phi}_0$ are defined by the same formula for $n = 0$).

13.5 Comparison of classical capacities

Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel from A to B, and let $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ be its complementary channel (see Definition 5.7.1 for the definition of a complementary channel), where $\mathbb{H}_E$ is the Hilbert space representing environment system E.

We have so far explored an unconstrained Holevo χ-capacity $C_\chi(\Phi)$, classical capacity $C(\Phi)$ and entanglement-assisted classical capacity $C_{\mathrm{ea}}(\Phi)$ for a memoryless channel Φ. The first of them, $C_\chi(\Phi)$, is defined as the maximum rate of information transmission between transmitter and receiver (generally called Alice and Bob) when nonentangled block coding is used by Alice and arbitrary measurement is used by Bob; the second one, $C(\Phi)$, differs from the first by the possibility of using arbitrary block coding by Alice; while the entanglement-assisted capacity, $C_{\mathrm{ea}}(\Phi)$ is defined as the maximum rate of information transmission between Alice and Bob under the assumption that they share a common entangled state, which can be used in block coding by Alice to increase the rate of information transmission.

By their operations definitions (see (13.7), (13.9) and (13.57)), we have $C_\chi(\Phi) \le C(\Phi) \le C_{\mathrm{ea}}(\Phi)$ in general. The main goal of this section is to make detailed comparisons among these three channel capacities and those of constrained versions $C_\chi(\Phi; \mathcal{A})$, $C(\Phi; \mathcal{A})$ and $C_{\mathrm{ea}}(\Phi; \mathcal{A})$, correspondingly, where $\mathcal{A}$ is a certain closed subset of $\mathcal{S}(\mathbb{H}_A)$.

13.5.1 Finite-dimensional unconstrained capacities

We make comparisons among the three unconstrained capacities $C_\chi(\Phi)$, $C(\Phi)$ and $C_{\mathrm{ea}}(\Phi)$ in finite-dimensional and infinite-dimensional cases below.

To draw some concrete conclusions, we first assume that $\mathbb{H}_A$, $\mathbb{H}_B$ and $\mathbb{H}_E$ are finite-dimensional Hilbert spaces. Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel, and $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ be its complementary channel, which can be uniquely defined up to unitary equivalence (see Definition 5.7.1). For the entanglement-assisted channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$, we assume that $A' = B$ for simplicity. Let $H(\rho)$ be the von Neumann entropy of $\rho \in \mathcal{S}(\mathbb{H}_A)$ and $H(\rho\|\sigma)$ be quantum relative entropy of ρ, relative to $\sigma \in \mathcal{S}(\mathbb{H}_A)$ (see, e. g., Lemma 8.2.5 for definition of relative entropy and approximation).

Recall that the Holevo χ-capacity $C_\chi(\Phi)$ is defined as (see Definition 12.2.1)

$$C_\chi(\Phi) = \max_{\rho \in \mathcal{S}(\mathbb{H}_A)} \chi_\Phi(\rho), \tag{13.101}$$

where

$$\chi_\Phi(\rho) = \max_{\sum_i p_i \rho_i = \rho} \sum_i p_i H(\Phi(\rho_i)\|\Phi(\rho)) \tag{13.102}$$

is the χ-function of Φ. Note that

$$\chi_{\Phi}(\rho) = H(\Phi(\rho)) - \hat{H}_{\Phi}(\rho), \tag{13.103}$$

where $\hat{H}_{\Phi}(\rho) = \min_{\sum_i p_i\rho_i=\rho} \sum_i p_i H(\rho_i)$ is the convex hull of the function $\rho \mapsto H(\Phi(\rho))$. By concavity of this function, the above minimum can be taken over ensembles of pure states. An ensemble $\{(p_i, \rho_i)\}$ is called optimal of the channel Φ (see Definition 12.2.5) if

$$C_\chi(\Phi) = \chi_{\Phi}(\bar{\rho}) = \sum_i p_i H(\Phi(\rho_i)\|\Phi(\bar{\rho})), \quad \bar{\rho} = \sum_i p_i \rho_i.$$

By the Holevo–Schumacher–Westmoreland theorem (see Theorem 13.2.2), the classical capacity of the channel Φ can be expressed by the following regularization formula:

$$C(\Phi) = \lim_{n\to+\infty} \frac{1}{n} C_\chi(\Phi^{\otimes n}).$$

By the Bennett–Shor–Smolin–Thapliyal theorem (see Theorem 13.4.1), the entanglement-assisted classical capacity of the channel Φ is determined as follows:

$$C_{\mathrm{ea}}(\Phi) = \max_{\rho\in\mathcal{S}(\mathbb{H}_A)} I_m(\rho;\Phi), \tag{13.104}$$

where $I_m(\rho;\Phi)$ is the mutual information of Φ at ρ defined by

$$\begin{aligned} I_m(\rho;\Phi) &= H(\rho) + H(\Phi(\rho)) - H(\hat{\Phi}(\rho)) \\ &= H(\rho) + \chi_{\Phi}(\rho) - \chi_{\hat{\Phi}}(\rho) \\ &= \chi_{\Phi}(\rho) + \Delta_{\Phi}(\rho), \end{aligned} \tag{13.105}$$

where $\Delta_{\Phi}(\rho) = H(\rho) - \chi_{\hat{\Phi}}(\rho)$. The above expression is obtained by using (13.103) and the fact that $\hat{H}_{\Phi}(\rho) = \hat{H}_{\hat{\Phi}}(\rho)$ because of the coincidence of the functions $\rho \mapsto H(\Phi(\rho))$ and $\rho \mapsto H(\hat{\Phi}(\rho))$ on the set of pure states.

Since $H(\rho) = \sum_i p_i H(\rho_i\|\rho)$ for any ensemble $\{(p_i, \rho_i)\}$ of pure states with average state ρ, we have

$$\Delta_{\Phi}(\rho) = \min_{\sum_i p_i\rho_i=\rho} \sum_i p_i\big(H(\rho_i\|\rho) - H(\hat{\Phi}(\rho_i)\|\hat{\Phi}(\rho))\big) \geq 0, \tag{13.106}$$

where ρ_i are rank one pure states and the last inequality follows from monotonicity of relative entropy, i. e., $H(\rho_i\|\rho) \geq H(\hat{\Phi}(\rho_i)\|\hat{\Phi}(\rho))$ (see Lemma 9.1.3).

The minimum in (13.106) is attained at an ensemble $\{(p_i, \rho_i)\}$ of pure states if and only if the maximum in (13.102) is attained at this ensemble. In addition, since $\sum_i p_i H(\Phi(\rho_i)) = \sum_i p_i H(\hat{\Phi}(\rho_i))$, this can easily shown by using expression (13.103) for χ-functions of the channels Φ and $\hat{\Phi}$.

Proposition 13.5.1. *Assume that* $\dim(\mathbb{H}_A) < +\infty$ *and* $\dim(\mathbb{H}_B) < +\infty$. *Then*

$$C_{\mathrm{ea}}(\Phi) \le C_\chi(\Phi) + \log(\dim(\mathbb{H}_A)) = C_\chi(\Phi) + nH(\rho). \tag{13.107}$$

Proof. This follows from the definition of $C_{\mathrm{ea}}(\Phi)$ for finite-dimensional $\mathbb{H}_A$. This proves the proposition. □

Corollary 13.5.2. *Assume that* $\dim(\mathbb{H}_A) < +\infty$ *and* $\dim(\mathbb{H}_B) < +\infty$. *The following inequalities hold:*

$$\begin{aligned} H(\rho_1) - C_\chi(\hat{\Phi}) \le C_{\mathrm{ea}}(\Phi) - C_\chi(\Phi) &\le H(\rho_2) - \chi_{\hat{\Phi}}(\rho_2) \\ &\le H(\Phi(\rho_2)) - \chi_{\hat{\Phi}}(\rho_2) = I_c(\Phi) + \hat{H}_\Phi(\rho_2), \end{aligned}$$

where ρ_1 *and* ρ_2 *are states in* $\mathcal{S}(\mathbb{H}_A)$ *such that* $\chi_\Phi(\rho_1) = C_\chi(\Phi)$ *and* $I_m(\rho_2;\Phi) = C_{\mathrm{ea}}(\Phi)$.

13.5.2 Infinite-dimensional constrained capacities

When dealing with infinite-dimensional quantum systems and channels, it is necessary to consider generalized ensembles defined as Borel probability measures μ on Borel subsets of $\mathcal{S}(\mathbb{H}_A)$. From this point of view, ordinary ensembles are described by finitely supported measures $\mu = \{(p_i, \rho_i)\}_{i=1}^N$, where N is a certain positive integer. We denote by $\mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ the set of all generalized ensembles of states in $\mathcal{S}(\mathbb{H}_A)$ (i. e., the space of all probability measures on the Borel measurable space $(\mathcal{S}(\mathbb{H}_A), \mathcal{B}(\mathcal{S}(\mathbb{H}_A))))$. The Holevo χ-quantity of a generalized ensemble μ is defined as

$$\chi(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} H(\rho \| \bar{\rho}(\mu))\mu(d\rho) = H(\bar{\rho}(\mu)) - \int_{\mathcal{S}(\mathbb{H}_A)} H(\rho)\mu(d\rho), \tag{13.108}$$

where the Bochner integral $\bar{\rho}(\mu) = \int_{\mathcal{S}(\mathbb{H}_A)} \rho\mu(d\rho)$ is the barycenter or average state of μ and the second formula is valid under the condition $H(\bar{\rho}(\mu)) < +\infty$ (see Proposition 12.1.4). Recall that the Holevo χ-quantity of the channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is defined by

$$\begin{aligned} \chi_\Phi(\mu) &= \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho) \| \Phi(\bar{\rho}(\mu)))\mu(d\rho) \\ &= H(\Phi(\bar{\rho}(\mu))) - \int_{\mathcal{S}(\mathbb{H}_A)} H(\Phi(\rho))\mu(d\rho) \end{aligned} \tag{13.109}$$

for all $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$, where the second equality is valid under the condition $H(\Phi(\bar{\rho}(\mu))) < +\infty$. It is shown in Proposition 12.1.4 that the constrained Holevo χ-capacity defined by Definition 12.2.1 can be expressed as

$$C_\chi(\Phi;\mathcal{A}) = \sup_{\mu\in\mathcal{P}(\mathcal{S}(\mathcal{A}))} \chi_\Phi(\mu) \tag{13.110}$$

for all closed subset $\mathcal{A}$ of $\mathcal{S}(\mathbb{H}_A)$ (the supremum is over $\mu \in \mathcal{P}(\mathcal{S}(\mathbb{H}_A))$ with the average state $\bar{\rho}(\mu) \in \mathcal{A}$). By letting $\mathcal{A} = \mathcal{K}_{\mathbf{H}}(E)$ for $\mathfrak{H}$-operator $\mathbf{H}$ on $\mathbb{H}_A$, we write $C_\chi(\Phi,\mathbf{H},E) = \sup_{\mu:\bar{\rho}(\mu)\in\mathcal{K}_{\mathbf{H}}(E)} \chi_\Phi(\mu)$.

In the following, we study general relations between the capacities $C_\chi(\Phi,\mathbf{H},E)$, $C(\Phi,\mathbf{H},E)$ and $C_{\mathrm{ea}}(\Phi,\mathbf{H},E)$ and give conditions for their coincidence under the assumption

$$H(\rho) < +\infty, \quad \forall\rho \in \mathcal{K}_{\mathbf{H}}(E), \tag{13.111}$$

which implies, in particular, finiteness of all these values. A basic role in this analysis is played by the following expression for the quantum mutual information:

$$I_m(\rho;\Phi) = H(\rho) + C_\chi(\Phi,\{\rho\}) - C_\chi(\hat{\Phi},\{\rho\}), \tag{13.112}$$

which is valid under the condition $H(\rho) < +\infty$ (since $C_\chi(\Phi,\{\rho\}) \le H(\rho)$ for any channel Φ). This condition implies finiteness of all terms on the right-hand side of (13.112).

If $H(\Phi(\rho)) = H(\hat{\Phi}(\rho))$ (this follows from the coincidence of $H(\Phi(\rho))$ and $H(\hat{\Phi}(\rho))$ for pure states ρ); in the general case one can show that this assumption holds if and only if $\mathrm{tr}[\exp(-\lambda\mathbf{H})] < +\infty$ for some $\lambda > 0$, which implies, in particular, finiteness for all these values.

By subadditivity of the quantum mutual information, expression (13.112) implies a formal proof of the inequality

$$C(\Phi;\mathbf{H},E) = \lim_{n\to+\infty} \frac{1}{n} C_\chi(\Phi;\mathbf{H},E) \le C_{\mathrm{ea}}(\Phi;\mathbf{H},E), \tag{13.113}$$

which looks obvious from the operational definitions of the capacities. It also implies the following inequalities.

Proposition 13.5.3. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel, and let $\mathbf{H}$ be an $\mathfrak{H}$-operator such that condition (13.111) is valid. The inequalities*

$$C_{\mathrm{ea}}(\Phi;\mathbf{H},E) \ge 2C_\chi(\Phi;\mathbf{H},E) - C_\chi(\hat{\Phi};\mathbf{H},E), \tag{13.114}$$

and

$$C_{\mathrm{ea}}(\Phi;\mathbf{H},E) \ge 2C(\Phi;\mathbf{H},E) - C(\hat{\Phi},\mathbf{H},E) \tag{13.115}$$

hold, where $\hat{\Phi} : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_E)$ is the complementary channel to Φ.

Note that, in contrast to (13.113), both inequalities (13.114) and (13.115) hold with equality if Φ is a noiseless channel. These inequalities show that coincidence of $C_{\mathrm{ea}}(\Phi;\mathbf{H},E) = C_\chi(\Phi;\mathbf{H},E)$.

Proof. For arbitrary $\epsilon > 0$, let ρ_ϵ be a state in $\mathcal{S}(\mathbb{H}_A)$ such that

$$C_\chi(\Phi; \mathbf{H}, E) < C_\chi(\Phi, \{\rho_\epsilon\}) + \epsilon, \quad \operatorname{tr}[\rho_\epsilon \mathbf{H}] \leq E.$$

Since $C_\chi(\Phi, \{\rho_\epsilon\}) \leq H(\rho_\epsilon) < +\infty$, Theorem 13.4.5 and formula (13.112) show that

$$\begin{aligned} C_{\mathrm{ea}}(\Phi; \mathbf{H}, E) \geq I_m(\rho_\epsilon; \Phi) &\geq 2C_\chi(\Phi, \{\rho_\epsilon\}) - C_\chi(\hat{\Phi}, \{\rho_\epsilon\}) \\ &\geq 2C_\chi(\Phi; \mathbf{H}, E) - C_\chi(\hat{\Phi}; \mathbf{H}, E) - 2\epsilon, \end{aligned}$$

which implies (13.114). Inequality (13.115) is obtained from (13.114) by regularization. This proves the proposition. □

14 Structure of quantum memory channels

In Chapter 13, we investigated n uses of memoryless quantum channels and studied their various properties and classical capacities associated with them. This chapter is devoted to the study of structures of memory quantum channels. As a special case, a class of memory channels, namely forgetful channels will also be explored. This chapter, however, does not address topics on the coding theorem and classical capacity. Special cases of quantum memory channels such as channels with Markovian and long-term memory and their classical capacities are the topics of interest in the next two chapters.

For each $n \in \mathbb{N}$, we use the block coding of length n and size N_n as considered in the previous chapter. Let the n-use of the channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ in transmitting time-ordered sequence of possibly entangled codewords $\rho^{(n)}_{\Lambda_n,A} \in \mathcal{S}(\mathbb{H}_A^{\otimes n})$ for classical information $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{N_n}\}$ be denoted by $\Phi^{\otimes n}(\rho^{(n)}_{\Lambda_n,A})$. Roughly speaking, the channel Φ is said to be a channel with memory or a memory channel or a channel with correlated noise, if $\Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A})$ cannot be expressed as n-fold tensor product $\Phi^{\otimes n}(\rho^{(n)}_{\Lambda_n,A})$, i. e.,

$$\Phi^{\otimes n}(\rho^{(n)}_{\Lambda_n,A}) \neq \Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A})$$

for some $\rho^{(n)}_{\Lambda_n,A}$.

For memoryless channels, the transmitted information and noise sources are treated as independent random variables, whereas for real-world noisy quantum channels, this independence assumption should be removed since the correlations between the errors are real and common. Examples of quantum channels, which naturally acquire a memory are common in quantum information processing. Recently, an unmodulated spin chain has been proposed as a model for short distance quantum communication. In such a scheme, the state to be communicated over the channel is placed in one end of the chain, propagates for a specific amount of time, and is then received at the other end. It is generally assumed that a reset of the spin chain occurs after each signal (resulting in a memoryless channel). However, a continuous operation without reset leads to higher transmission rates and corresponds to quantum memory channel.

The first model of a quantum memory channel was studied by Macchiavello and Palma [111]. They showed that the transmission of classical information through two successive uses of a quantum depolarizing channel, with Markovian correlated noise, is enhanced by using inputs entangled over the two uses. In order to describe memory channels, Kretschmann and Werner [100] proposed, in addition to Alice's input register system $\mathbb{H}_A$ and Bob's output register system $\mathbb{H}_B$, that the transmission process also involves an additional memory input $\mathbb{H}_M$ and an additional finite-dimensional memory output $\mathbb{H}_{M'}$ be also taken into consideration in the formulation of such channels. Readers are also referred to the review paper on the quantum channel with memory

https://doi.org/10.1515/9783110788105-014

effects by Caruso et al. [18] for a thorough survey on various aspects of memory quantum channels.

14.1 Representations of memoryless channels

In this section, Kraus and unitary representations of n-use of memoryless channel $\Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A}) = \Phi^{\otimes n}(\rho^{(n)}_{\Lambda_n,A})$ in both Schrodinger and Heisenberg pictures are represented as a motivation for describing these two pictures in terms of their memory counterparts, which are to be studied in detail in Subsections 14.2.2, 14.2.1 and 14.2.3.

The presentation of this section is largely based on results obtained in Giovannetti [53] (see also Caruso et al. [18]).

14.1.1 Kraus representation in the Schrodinger picture

Consider a quantum channel Φ with a Kraus representation (see Theorem 4.4.4)

$$\Phi(\rho) = \sum_{i=1}^{+\infty} \mathbf{K}_i \rho \mathbf{K}_i^*, \quad \forall \rho \in \mathcal{S}(\mathbb{H}_A),$$

where $\mathbf{K}_i$ is a *-weakly continuous linear map from $\mathbb{H}_A$ to $\mathbb{H}_B$ such that

$$\sum_{n=1}^{+\infty} \mathbf{K}_i \mathbf{K}_i^*$$

converges strongly to the identity operator $\mathbf{I}_B$ (the identity operator in $\mathbb{H}_B$) in $\|\cdot\|_1$-norm.

As described in Kretschmann and Werner [100], the action of a quantum channel Φ on the input state density operator $\rho^{(n)}_{\Lambda_n,A} \in \mathcal{S}(\mathbb{H}^{\otimes n}_{\Lambda_n,A})$, consisting of n qubits (including entangled ones) is given by

$$\Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A}) = \sum_{k_1,k_2,\dots,k_n} p_{k_1\cdots k_n}(\mathbf{K}_{k_n} \otimes \cdots \otimes \mathbf{K}_{k_1})\rho^{(n)}_{\Lambda_n,A}(\mathbf{K}^*_{k_1} \otimes \cdots \otimes \mathbf{K}^*_{k_n}) \tag{14.1}$$

where the Kraus operators $\mathbf{K}_{k_n} \otimes \cdots \otimes \mathbf{K}_{k_1}$ are applied with probability $p_{k_1,\dots,k_n}$, which satisfies $\sum_{k_1,\dots,k_n} p_{k_1,\dots,k_n} = 1$. The quantity $p_{k_1,\dots,k_n}$ can be interpreted as the probability that a random sequence of operations is applied to the sequence of n qubits transmitted through the channel. For a memoryless channel, these operations are independent, therefore, $p_{k_1,k_2,\dots,k_n} = p_{k_1}p_{k_2}\cdots p_{k_n}$. That is,

$$\begin{aligned}\Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A}) &= \Phi^{\otimes n}(\rho^{(n)}_{\Lambda_n,A}) \\ &= \sum_{i_1,i_2,\dots,i_n} p_{i_1}p_{i_2}\cdots p_{i_n}(\mathbf{K}_{i_n} \otimes \cdots \otimes \mathbf{K}_{i_1})\rho^{(n)}_{\Lambda_n,A}(\mathbf{K}^*_{i_1} \otimes \cdots \otimes \mathbf{K}^*_{i_n}),\end{aligned}$$

for all $\rho^{(n)}_{\Lambda_n,A} \in \mathcal{S}(\mathbb{H}^{\otimes n}_A)$.

From the above, one can say that the Kraus operators of the memoryless channel $\Phi^{(n)}$ can be expressed as a tensor product $\mathbf{K}_{i_1} \otimes \cdots \otimes \mathbf{K}_{i_n}$ formed by independent and identically distributed sequences extracted from the Kraus set $\{\mathbf{K}_i\}_{i=1}^{+\infty}$ associated with the single carrier channel Φ. In the presence of memory, they exhibit some correlation. A simple example is given by the Markov chain, i. e.,

$$p_{k_1,\ldots,k_n} = p_{k_1} p_{k_2|k_1} \cdots p_{k_n|k_{n-1}}. \tag{14.2}$$

In the above expression, $p_{k_n|k_{n-1}}$ is the conditional probability that an operation, say $\mathbf{K}_{k_n}$, is applied to the nth qubit provided that it was applied on the $(n-1)$-th qubit.

14.1.2 Kraus representation in the Heisenberg picture

Kraus' representation of the dual memoryless channel $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \to \mathfrak{B}(\mathbb{H}_A)$ in the Heisenberg picture can be expressed as

$$\Phi^*(\mathbf{A}) = \sum_{i=1}^{+\infty} \mathbf{K}_i^* \mathbf{A} \mathbf{K}_i, \quad \forall \mathbf{A} \in \mathfrak{B}(\mathbb{H}_B). \tag{14.3}$$

The above expression is obtained via the following duality relation:

$$\begin{aligned} \operatorname{tr}[\mathbf{A}\Phi(\rho)] &= \operatorname{tr}\left[\mathbf{A} \sum_{i=1}^{+\infty} \mathbf{K}_i \rho \mathbf{K}_i^*\right] = \sum_{i=1}^{+\infty} \operatorname{tr}[\mathbf{A}\mathbf{K}_i \rho \mathbf{K}_i^*] \\ &= \sum_{i=1}^{+\infty} \operatorname{tr}[\mathbf{K}_i^* \mathbf{A} \mathbf{K}_i \rho] = \operatorname{tr}\left[\sum_{i=1}^{+\infty} \mathbf{K}_i^* \mathbf{A} \mathbf{K}_i \rho\right] = \operatorname{tr}[\Phi^*(\mathbf{A})\rho]. \end{aligned} \tag{14.4}$$

Note that the dual of a channel needs not be a channel, that is, it might not be trace preserving. Instead, the completeness property of the Kraus operators $\{\mathbf{K}_i\}_{i=1}^{+\infty}$ implies that $\Phi^*(\mathbf{I}_B) = \mathbf{I}_A$. By the same token, the Kraus representation of $\Phi^{(n)} : \mathfrak{B}(\mathbb{H}_B^{\otimes n}) \to \mathfrak{B}(\mathbb{H}_A^{\otimes n})$, the dual of memory channel $\Phi^{(n)}$, can be written as

$$\begin{aligned} &\Phi^{*(n)}(\mathbf{A}^{(n)}) \\ &\quad = \sum_{k_1,k_2,\ldots,k_n} p_{k_1,k_2,\ldots,k_n} (\mathbf{K}_{i_n}^* \otimes \cdots \otimes \mathbf{K}_{i_1}^*) \mathbf{A}^{(n)} (\mathbf{K}_{i_1} \otimes \cdots \otimes \mathbf{K}_{i_n}), \end{aligned} \tag{14.5}$$

for all $\mathbf{A}^{(n)} : \mathfrak{B}(\mathbb{H}_B^{\otimes n}) \to \mathfrak{B}(\mathbb{H}_A^{\otimes n})$.

14.1.3 Unitary representation in the Schrodinger picture

Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel from the sender Alice to the receiver Bob. By the Stinespring dilation theorem (see Theorem 4.3.4), any such map may be represented as a unitary operation between the input state ρ_A and a known environment state ω_E. In this case, the output state of a single use of the channel is given by

$$\rho_B = \Phi(\rho_A) = \mathrm{tr}_E[\mathbf{U}_{AE}(\rho_A \otimes \omega_E)\mathbf{U}_{AE}^*] \tag{14.6}$$

with $\rho_A \in \mathcal{S}(\mathbb{H}_A)$ being the state sent by Alice, and $\rho_B \in \mathcal{S}(\mathbb{H}_B)$ being the output state received by Bob, where $\mathbf{U}_{AE}$ is an unitary operator on $\mathcal{S}(\mathbb{H}_{AE})$. For a sequence of transmissions of codewords for $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{N_n}\}$ through the channel, we assume that transmission of $\lambda \in \Lambda_n$ interacts with independent and identical environment state ω_n for $n = 1, 2, \ldots$. We write $\omega_E^{(n)} = \omega_1 \otimes \omega_2 \otimes \cdots \otimes \omega_n$. In this case, the memoryless output of n transmissions can be written as

$$\rho_{\Lambda_n,B}^{(n)} = \Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)}) = \Phi^{\otimes n}(\rho_{\Lambda_n,A}^{(n)}), \tag{14.7}$$

where

$$\Phi^{\otimes n}(\rho_{\Lambda_n,A}^{(n)}) = \mathrm{tr}_E[\mathbf{U}_{\lambda_n,E_n} \cdots \mathbf{U}_{\lambda_1,E_1}(\rho_{\Lambda_n,A}^{(n)} \otimes \omega_E^{(n)})\mathbf{U}_{\lambda_1,E_1}^* \cdots \mathbf{U}_{\lambda_n,E_n}^*]. \tag{14.8}$$

In the above, the state $\rho_{\Lambda_n,A}^{(n)}$ now represents a (possibly entangled) input state across the n channel uses. The unitary operations $\mathbf{U}_{\lambda_k,E_k}$ are all identical, and the environment state is a product state $\omega_E^{(n)} = |0\rangle_{E_1}\langle 0| \otimes \cdots \otimes |0\rangle_{E_n}\langle 0|$. Thus, we may write $\Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)}) = \Phi^{\otimes n}(\rho_{\Lambda_n,A}^{(n)})$.

The presentation we made above of quantum channels and n-uses of these memoryless quantum channels is typical of the Schrodinger picture of quantum mechanics. The Schrodinger picture can be seen as being the resulting transformation of a quantum state of $\mathbb{H}_A$ after a contact and an evolution with some environment. As usual, for all quantum evolutions there is a dual picture, an Heisenberg picture, where the evolution is seen from the point of view of observables instead of states. It so happens that in the case of quantum channels this dual picture opens the door to a vast and interesting field: the notion of completely positive maps that we shall explore in the next section.

14.2 Constructive approach to memory channels

Bowen and Werner [14] (see also Caruso et al. [18]) proposed a constructive approach to memory channels described below.

For each $n \in \mathbb{N}$, let $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{N_n}\}$ be a collection of classical information to be encoded and sent sequently by Alice through the noisy quantum memory channel Φ to Bob. In the following, we describe a model for transmitting a codeword down through a memory channel Φ. Here, one uses transmission of the code word through the memory channel and can be written as a completely positive trace preserving map

$$\Phi : \mathcal{S}(\mathbb{H}_{AM}) \to \mathcal{S}(\mathbb{H}_{BM}),$$

where $\mathbb{H}_M$ is a Hilbert space that represents memory introduced by Bowen and Mancini [14] and Kretschmann and Werner [100] and $\mathbb{H}_{AM} := \mathbb{H}_A \otimes \mathbb{H}_M$, $\mathbb{H}_{BM} := \mathbb{H}_B \otimes \mathbb{H}_M$.

If Alice transmits an encoded input state $\rho \in \mathcal{S}(\mathbb{H}_A)$ with the initial memory state $|0\rangle_M\langle 0| \in \mathcal{S}(\mathbb{H}_M)$, then Bob will receive the output state $\Phi(\rho \otimes (|0\rangle_M\langle 0|))$. If Bob is not interested in the information on its memory component (or cannot access it), he can ignore the memory output by tracing away $\mathbb{H}_M$ such that the real output state received by Bob is $\mathrm{tr}_M[\Phi(\rho \otimes |0\rangle_M\langle 0|)]$. To send an n-fold system in the state $\rho^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n})$, we have to invoke the memory channel Φ n times in succession. The kth invocation is again a tensor product, but now the memory has to be taken into account such that we get

$$\Phi_k = \underbrace{\mathfrak{I} \otimes \cdots \otimes \mathfrak{I}}_{k-1} \otimes \Phi \otimes \underbrace{\mathfrak{I} \otimes \cdots \otimes \mathfrak{I}}_{n-k}, \tag{14.9}$$

which is a map of the form

$$\Phi_k : \mathfrak{B}(\mathbb{H}_A^{\otimes(k-1)} \otimes \mathbb{H}_M \otimes \mathbb{H}_A^{\otimes(n-k)}) \to \mathfrak{B}(\mathbb{H}_A^{\otimes k} \otimes \mathbb{H}_M \otimes \mathbb{H}_B^{\otimes(n-k-1)}). \tag{14.10}$$

Note that the factor $\mathbb{H}_M$ is shifted here from kth to $(k+1)$-th position. This allows us to write the overall operation as in (13.2) as a concatenation

$$\Phi^{(n)} : \mathfrak{B}(\mathbb{H}_M \otimes \mathbb{H}_A^{\otimes n}) \to \mathfrak{B}(\mathbb{H}_B^{\otimes n} \otimes \mathbb{H}_M) \tag{14.11}$$

$$\Phi^{(n)} = \Phi_n \circ \Phi_{n-1} \circ \cdots \circ \Phi_1. \tag{14.12}$$

If the memory is ignored at the end, Bob received the n-fold system as above in the final state $\mathrm{tr}_M[\Phi^{(n)}(\rho \otimes |0\rangle_M\langle 0|)]$. Note that in the contrast to the memoryless $\Phi^{(n)}$ we cannot write $\Phi^{(n)}$ as a tensor product $\Phi^{\otimes n}$ and even if the state ρ is a product state the output state is not in general.

The scheme just constructed above describes a channel that can act on an arbitrary number n of systems (via the concatenations $\Phi^{(n)}$). Furthermore, it satisfies the natural causality condition that the kth invocation depends on the $k-1$ previous ones but not the $n-k$ that will take place in the future. It can be shown that is causal in this way can be written as a concatenation of a memory channel Φ.

The dimension of the memory is determined by the number of Kraus operators in the single channel expansion and the correlation length of the channel, which may be defined as the maximum number of channel uses for which the noise is not conditionally independent. Any channel with a finite correlation length may be generated by a channel with a finite memory, according to this model.

Bowen and Mancini [14] proposed a constructive approach to modeling quantum memory channels described below.

Definition 14.2.1 (Consistency condition). A family of CPTP maps $\{\Phi^{(n)}\}_{n=1}^{+\infty}$, $\Phi^{(n)} : \mathcal{S}(\mathbb{H}_A^{(n)}) \to \mathcal{S}(\mathbb{H}_B^{(n)})$, is said to be consistent if for all $1 \le m < n$ and for all $\rho_{\Lambda_n,A}^{(n)} \in \mathcal{S}(\mathbb{H}_A^{(n)})$,

$$\Phi^{(m)}(tr^{(m)}[\rho_{\Lambda_n,A}^{(n)}]) = tr^{(m)}[\Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)})], \tag{14.13}$$

where $\mathrm{tr}^{(m)}[\rho_{\Lambda_n,A}^{(n)}]$ denotes the partial trace over all carriers but the first m and $\mathrm{tr}^{(m)}[\Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)})]$ denotes the partial trace over all received carriers but the first m.

In memory quantum channels to be discussed below, we shall treat those models in which the noise respects the time-ordering of the carriers for the classical information Λ so that at a given channel use, the output cannot be influenced by successive inputs. This property generalizes the notion of semicausality discussed in Subsection 14.3.2 to the case of multiple (ordered) subsystems. Inspired by the classical theory of communication (Gallager [55]), one can name the quantum communication lines, which fulfills such condition, nonanticipatory quantum channels. Notice, however, that in the approach of Kretschmann and Werner [100] these maps are called just causal. Under the nonanticipatory condition, there must exist a family of CPTP maps $\{\Phi^{(n)}\}_{n=1}^{+\infty}$ with $\Phi^{(n)} : \mathcal{S}(\mathbb{H}_A^{(n)}) \to \mathcal{S}(\mathbb{H}_B^{(n)})$, which allows one to express the output states of the first n carriers in terms of the density operators of their associated inputs. Precisely, the definition of nonanticipatory condition is given below.

Definition 14.2.2 (Nonanticipatory condition). A consistent family of CPTP maps $\{\Phi^{(n)}\}_{n=1}^{+\infty}$, $\Phi^{(n)} : \mathcal{S}(\mathbb{H}_A^{(n)}) \to \mathcal{S}(\mathbb{H}_B^{(n)})$ is said to be nonanticipatory if the output states of the first n carriers in terms of the quantum states of their associated inputs, i. e.,

$$\rho_{\Lambda_n,A}^{(n)} \mapsto \Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)}),$$

does not depend on $\rho_{\Lambda_m,A}^{(m)}$ for $m > n$.

A formal definition of quantum memory channels is given below.

Definition 14.2.3 (Definition of memory quantum channels). Let $(\Phi^{(n)})_{n=1}^{+\infty}$ be a sequence of CPTP maps, where $\Phi^{(n)} : \mathcal{S}(\mathbb{H}_A^{(n)}) \to \mathcal{S}(\mathbb{H}_B^{(n)})$ represents the n-use of the quantum channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$. Then the quantum channel Φ is said to be a quantum memory channel if the sequence $(\Phi^{(n)})_{n=1}^{+\infty}$ satisfies the following three conditions:

1. The family satisfies the consistency condition defined by Definition 14.2.1;
2. The family satisfies the nonanticipatory condition defined by Definition 14.2.2;
3. $\Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)})$ cannot be written as $\Phi^{\otimes n}(\rho_{\Lambda_n,A}^{(n)})$. That is,

$$\Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)}) \ne \Phi^{\otimes n}(\rho_{\Lambda_n,A}^{(n)}) \quad \text{for some } \rho_{\Lambda_n,A}^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n})$$

where $\Phi^{\otimes n}(\rho_{\Lambda_n,A}^{(n)}) = (\Phi \otimes \cdots \otimes \Phi)(\rho_{\Lambda_n,A}^{(n)})$ for all n.

Some special cases of quantum memory channels such as channels with Markov memory will be treated in the following chapters.

14.2.1 Unitary representation in the Schrodinger picture

In order to mathematically describe memory channels, Kretschmann and Werner [100] and Bowen and Mancini [14] proposed that memory channels be modeled as a unitary interaction between the states transmitted through the channel, independent environment and the channel memory state that remains unchanged during the interaction. That is, in addition to Alice's input register system $\mathbb{H}_A$ and Bob's output register system $\mathbb{H}_B$, the transmission process also involves an additional memory input $\mathbb{H}_M$ and an additional finite-dimensional memory output $\mathbb{H}_{M'}$ be also taken into consideration in the formulation of such channels, since the smaller of the two Hilbert spaces $\mathbb{H}_M$ and $\mathbb{H}_{M'}$ can always be thought of as being embedded in the larger one. In the following, we will assume without loss of generality that $\mathbb{H}_M = \mathbb{H}_{M'}$ and $\dim(\mathbb{H}_M) < +\infty$.

In this model of a quantum memory channel, the input state $\rho_{\lambda,A} \in \mathcal{S}(\mathbb{H}_A)$ going through the channel interacts unitarily with an identical channel memory state ω_M, as well as an independent environment E_i for $i = 1, 2, \ldots$. Therefore, after one use of the channel the output state $\rho_{\lambda,B}$ after sending the codeword for classical data λ can be written as

$$\rho_{\lambda,B} = \mathbf{U}_{\lambda,M,E_i}(\rho_{\lambda,A} \otimes \omega_M \otimes \omega_{E_i})\mathbf{U}^*_{\lambda,M,E_i},$$

where $\mathbf{U}_{\lambda,M,E_i}$ is an unitary operator on the space $\mathcal{S}(\mathbb{H}_A \otimes \mathbb{H}_M \otimes \mathbb{H}_{E_i})$ and $\mathbf{U}^*_{\lambda,M,E_i}$ is the adjoint operator $\mathbf{U}_{\lambda,M,E_i}$. The backaction of the channel state on the message state therefore gives a memory to the channel. The general model thus includes a channel memory M, and the independent environments for each qubit E_i. Memory of a quantum channel can be described as subspace of the environment, which does not evolve over the time scale of its successive uses. The dimension of memory subspace determines the number of channel uses for which the noise is not conditionally independent.

Hence, the resulting map in a unitary representation (in the Schrodinger picture) of n-uses of a memory channel can be written as

$$\begin{aligned}\rho^{(n)}_{\Lambda_n,B} &= \Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A}) \\ &= \mathrm{tr}_{ME}[\mathbf{U}_{n,M,E_n}\cdots\mathbf{U}_{1,M,E_1}(\rho^{(n)}_{\Lambda_n,A} \otimes \omega_M \otimes \omega^{(n)}_E)\mathbf{U}^*_{1,M,E_1}\cdots\mathbf{U}^*_{n,M,E_n}], \qquad (14.14)\end{aligned}$$

where $\mathrm{tr}_{ME}[\cdots]$ is the partial trace taken with respect to $\mathbb{H}_{ME} := \mathbb{H}_M \otimes \mathbb{H}_E$.

If the unitaries factor into independent unitaries acting on the memory and the combined state and environment, that is, $\mathbf{U}_{n,M,E_n} = \mathbf{U}_{n,E_n}\mathbf{U}_M$, then the memory traces out and we have a memoryless channel. If the unitaries reduce to $\mathbf{U}_{n,M}$, we can call it a perfect memory channel, as no information is lost to the environment.

Example 14.1. Consider the Kraus representation of a Pauli channel defined by

$$\rho \mapsto \Phi(\rho) = \sum_{i=0}^{3} \mathbf{K}_i \rho \mathbf{K}_i^* \tag{14.15}$$

where $\mathbf{K}_i = \sqrt{p_i}\sigma_i$, where σ_i, $i = 0, 1, 2, 3$ are defined by the Pauli matrices

$$I = \sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -\imath \\ \imath & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Here, we consider two consecutive uses of a Pauli channel with partial memory (see Marchiavello and Palma [111]), i. e., $p_{k_n|k_{n-1}} = (1-\mu)p_{k_n} + \mu\delta_{k_n|k_{n-1}}$. This means that with probability μ the same rotation is applied to both qubits, while with probability $1-\mu$, the two rotations are uncorrelated. Then

$$\Phi^{(2)}(\rho^{(2)}) = \sum_{i,j=0}^{3} K_{i,j}\rho^{(2)}\mathbf{K}_{i,j}^*$$

where the Kraus operators can be expressed as

$$\mathbf{K}_{i,j} = \sqrt{p_i[(1-\mu)p_j + \mu\delta_{i,j}]}\sigma_i \otimes \sigma_j, \quad 0 \le \mu \le 1, \tag{14.16}$$

where μ is the memory coefficient of the channel and σ_i, $i, j = 0, 1, 2, 3$ are the Pauli operators with $\sigma_0 = I$.

14.2.2 Unitary representation in the Heisenberg picture

To illustrate an unitary representation for a quantum memory channel, we assume long messages with $n \in \mathbb{N}$ signal states are to be processed by subsequent application of memory channels Φ, resulting in the *concatenated channel* $\Phi^{*(n)} : (\mathfrak{B}(\mathbb{H}_B))^{\otimes n} \otimes \mathfrak{B}(\mathbb{H}_M) \to \mathfrak{B}(\mathbb{H}_M) \otimes (\mathfrak{B}(\mathbb{H}_A))^{\otimes n}$ given as follows:

$$\Phi^{*(n)} = (\Phi^* \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes(n-1)}) \circ \cdots \circ (\mathfrak{I}_{\mathcal{B}}^{\otimes(n-2)} \otimes \Phi^* \otimes \mathfrak{I}_{\mathcal{A}}) \circ (\mathfrak{I}_{\mathcal{B}}^{\otimes(n-1)} \otimes \Phi^*), \tag{14.17}$$

where $\mathfrak{I}_{\mathcal{A}}$ and $\mathfrak{I}_{\mathcal{B}}$ denote the identity operator on $\mathcal{A} = \mathfrak{B}(\mathbb{H}_A)$ and $\mathcal{B} = \mathfrak{B}(\mathbb{H}_B)$, respectively.

The above two different representations in the Schrodinger picture (14.14) and in the Heisenberg picture (14.17) of the n-use of the memory channel $\Phi^{(n)}$ and its dual channel $\Phi^{*(n)}$ are actually equivalent in the sense that

$$\mathrm{tr}[\Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)})(\cdots)] = \mathrm{tr}[\rho_{\Lambda_n,A}^{(n)}\Phi^{*(n)}(\cdots)]. \tag{14.18}$$

14.2.3 Kraus representation of the memory channel

If the state input to the memory channel Φ is $\rho_{\lambda,A} \in \mathcal{S}(\mathbb{H}_A)$, then the action of the channel is described as

$$\Phi(\rho_{\lambda,A}) = \sum_{k=1}^{+\infty} \mathbf{K}_k \rho_{\lambda,A} \mathbf{K}_k^*$$

where $\{\mathbf{K}_k\}_{k=1}^{+\infty}$ are the Kraus operators with $\sum_{k=1}^{+\infty} \mathbf{K}_k^* \mathbf{K}_k = \mathbf{I}$.

The action of the memory channel Φ on the input state $\rho_{\Lambda_n,A}^{(n)}$, consisting of n qubits (including entangled ones) is given by

$$\Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)}) = \sum_{k_1,k_2,\ldots,k_n} p_{k_1} \cdots p_{k_n} (\mathbf{K}_{k_n} \otimes \cdots \otimes \mathbf{K}_{k_1} \rho_{\Lambda_n,A}^{(n)} \mathbf{K}_{k_n}^* \otimes \cdots \otimes \mathbf{K}_{k_n}^*), \tag{14.19}$$

where the Kraus operators $\mathbf{K}_{k_n} \cdots \mathbf{K}_{k_1}$ are applied with probability $p_{k_1,k_2,\ldots,k_n} > 0$, which satisfies $\sum_{k_1,k_2,\ldots,k_n} p_{k_1,k_2,\ldots,k_n} = 1$. The quantity $p_{k_1,\ldots,k_n} > 0$ can be interpreted as the probability that a random sequence of operations is applied to the sequence of n qubits transmitted through the channel.

We have the following special cases:

(i) If $p_{k_1,k_2,\ldots,k_n} = p_{k_1} p_{k_2} \cdots , p_{k_n}$, then Φ is a memoryless channel, which was treated in Chapter 13; and

(ii) If $p_{k_1,k_2,\ldots,k_n} = p_{k_1} p_{k_1,k_2} \cdots p_{k_{n-1},k_n}$, then (14.19) indicates that p_{k_{n-1},k_n} is the conditional probability that an operation, say $\mathbf{K}_{k_n}$, is applied to the nth qubit provided that it was applied on the $(n-1)$th qubit. In this case, Φ is a channel with Markovian memory, which will be treated in the next chapter.

14.3 Quasilocal algebras approach

Until now, we have followed a constructive approach in which memory quantum channels were always thought of as concatenations of smaller units which, starting from an official "first carrier" element, process one quantum signal each. An alternative view, where the communication lines are treated as mappings applied on infinitely long message strings, is proposed in Bjelakovic and Boche, [12] and Kretschmann and Werner [100] (see also the Appendix in Caruso et al. [18]).

This approach requires some advanced mathematical tools that are briefly reviewed in the following subsection.

14.3.1 Construction of quasilocal algebras

Quasilocal algebras are the proper mathematical tools to describe infinitely extended quantum lattice systems (Brattelli and Robinson [15]). For the sake of simplicity, let us consider a chain of infinitely many qubits (spins) placed in a one-dimensional lattice $\mathbb{Z}$. Then, to each lattice site $j \in \mathbb{Z} = \{\dots -3, -2, -1, 0, 1, 2, \dots\}$, attach a Hilbert space $\mathbb{H}_j$ and consider the associated von Neumann algebra of bounded linear operators $\mathcal{A}_j = \mathfrak{B}(\mathbb{H}_j)$ on the quantum subsystem $\mathbb{H}_j$ located at the site j.

When $\mathbb{A} \subset \mathbb{Z}$, we consider the following two cases:

(A) If $\mathbb{A} \subset \mathbb{Z}$ is a finite subset, we denote $\mathbb{H}_{\mathbb{A}}$, the chain of quantum subsystems located at sites $\mathbb{A}$ and $\mathcal{A}_{\mathbb{A}}$, the algebra of observables belonging to the sites $\mathbb{A}$ by

$$\mathbb{H}_{\mathbb{A}} = \bigotimes_{j \in \mathbb{A}} \mathbb{H}_j \quad \text{and} \quad \mathcal{A}_{\mathbb{A}} = \bigotimes_{j \in \mathbb{A}} \mathcal{A}_j. \tag{14.20}$$

Operators $\mathbf{a} \in \mathcal{A}_{\mathbb{A}}$ are called *local operators* as they are operators "localized" in $\mathbb{A}$.

Whenever $\mathbb{A}_1 \subset \mathbb{A}_2 \subset \mathbb{Z}$, a localized operator $\mathbf{a} \in \mathcal{A}_{\mathbb{A}_1}$ can be extended to $\mathcal{A}_{\mathbb{A}_2}$ by tensoring with the identity operator on $\mathbb{A}_2 \backslash \mathbb{A}_1$. That is,

$$\mathbf{a} \in \mathcal{A}_{\mathbb{A}_1} \quad \text{and} \quad \mathbf{a} \otimes \mathbf{I}_{\mathbb{A}_2 \backslash \mathbb{A}_1} \in \mathcal{A}_{\mathbb{A}_2}. \tag{14.21}$$

The above two operators describe the same physical object. Therefore, we can identify $\mathcal{A}_{\mathbb{A}_1}$ with the subalgebra $\mathcal{A}_{\mathbb{A}_1} \otimes \mathbf{I}_{\mathbb{A}_2 \backslash \mathbb{A}_1}$ of $\mathcal{A}_{\mathbb{A}_2}$ through the map

$$\mathcal{A}_{\mathbb{A}_1} \ni \mathbf{a} \mapsto \mathbf{a} \otimes \mathbf{I}_{\mathbb{A}_2 \backslash \mathbb{A}_1} \in \mathcal{A}_{\mathbb{A}_2}. \tag{14.22}$$

To construct *-algebra of observables on the local quantum chain of the lattice system $\mathcal{A}^{\mathrm{loc}} := \bigcup_{\mathbb{A} \subset \mathbb{Z}} \mathcal{A}_{\mathbb{A}}$ (where the union above is taken for all finite subsets $\mathbb{A}$ of $\mathbb{Z}$), we define the addition "+," operator multiplication "$\circ$," adjointness "$*$" and the norm "$\|\cdot\|$" of observables on $\mathcal{A}^{\mathrm{loc}}$ as follows.

(i) For $\mathbf{a}_i \in \mathcal{A}_{\mathbb{A}_i}$ and $c_i \in \mathbb{C}$ for $i = 1, 2$, we define $c_1\mathbf{a}_1 + c_2\mathbf{a}_2$ by the above extension as

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 \mapsto c_1(\mathbf{a}_1 \otimes \mathbf{I}_{(\mathbb{A}_1 \cup \mathbb{A}_2)\backslash \mathbb{A}_1}) + c_2(\mathbf{a}_2 \otimes \mathbf{I}_{(\mathbb{A}_1 \cup \mathbb{A}_2)\backslash \mathbb{A}_2}).$$

(ii) In the same way, the product $\mathbf{a}_1\mathbf{a}_2 := \mathbf{a}_1 \circ \mathbf{a}_2$ of operators $\mathbf{a}_i \in \mathbb{A}_i$, $i = 1, 2$ can be extended to $\mathbb{A}_1 \cup \mathbb{A}_2$ by

$$\mathbf{a}_1\mathbf{a}_2 \mapsto (\mathbf{a}_1 \otimes \mathbf{I}_{(\mathbb{A}_1 \cup \mathbb{A}_2)\backslash \mathbb{A}_1})(\mathbf{a}_2 \otimes \mathbf{I}_{(\Lambda_1 \cup \Lambda_2)\backslash \Lambda_2}).$$

(iii) The adjoints $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$ of $\mathbf{a}_1$ and $\mathbf{a}_2$, respectively, are extended to the common set of sites $\Lambda_1 \cup \Lambda_2$.

(iv) Since tensoring with the identity operator $\mathbf{I}_{\mathcal{A}}$ does not change the norm, for $\mathbf{a}_i \in \mathcal{A}_{\Lambda_i}$, $i = 1, 2$, we have

$$\|\mathbf{a}_i\| = \|\mathbf{a}_i \otimes \mathbf{I}_{(\Lambda_1 \cup \Lambda_2)\backslash \Lambda_i}\|.$$

We define

$$\mathcal{A}^{\mathrm{loc}} = \bigcup_{\text{finite } \mathbb{A} \subset \mathbb{Z}} \mathcal{A}_{\mathbb{A}}. \tag{14.23}$$

Note that $\mathcal{A}^{\mathrm{loc}}$ is not complete under the norm $\| \cdot \|$ defined in (iv) above, and we consider the C^*-algebra $\mathcal{A}_{\mathbb{Z}}$ defined by (14.23) under the norm that yields a normed algebra of *local observables*, and will be denoted by

$$\mathcal{A}_{\mathbb{Z}} = \overline{\mathcal{A}^{\mathrm{loc}}}^{\|\cdot\|}, \tag{14.24}$$

where $\overline{\mathcal{A}^{\mathrm{loc}}}^{\|\cdot\|}$ denotes the closure of $\mathcal{A}^{\mathrm{loc}}$ under the norm $\| \cdot \|$.

(B) For infinite subset $\mathbb{A} \subset \mathbb{Z}$, we define $\mathcal{A}_{\mathbb{A}}$ as the closure of the union of all $\mathcal{A}_{\mathbb{A}'}$ for finite $\mathbb{A}' \subset \mathbb{A}$. That is,

$$\mathcal{A}_{\mathbb{A}} = \overline{\bigcup_{\mathbb{A}' \subset \mathbb{A} \text{ finite}} \mathcal{A}_{\mathbb{A}'}}^{\|\cdot\|}. \tag{14.25}$$

In particular, we will consider the left-half and right-half chains $\mathcal{A}_{]-\infty,0]}$ and $\mathcal{A}_{[1,\infty[}$, respectively.

The algebra $\mathcal{A}_{\mathbb{A}}$ is interpreted as the algebra of physical observables for a subsystem localized in the region $\mathbb{A} \subset \mathbb{Z}$. The *quasilocal algebra* $\mathcal{A}_{\mathbb{Z}}$ then corresponds to the extended algebra of observables on the infinite chain $\mathbb{Z}$.

Here, quasilocal stands for the fact that $\mathcal{A}_{\mathbb{Z}}$ besides all local observables, it also contains nonlocal observables, which can be approximated in norm by local ones. $\mathcal{A}_{\mathbb{Z}}$ results a C^*-algebra and its elements can be regarded in many respects as bounded operators.

It is also useful to consider methods to transform abstract elements $\mathbf{a} \in \mathcal{A}_{\mathbb{Z}}$ into operators $\pi(\mathbf{a})$ acting on a Hilbert space $\mathbb{H}$, i. e., representations of quasilocal algebra. A representation π on the Hilbert space $\mathbb{H}$ is a homomorphism $\pi : \mathcal{A}_{\mathbb{Z}} \to \mathfrak{B}(\mathbb{H})$, i. e., a linear map satisfying $\pi(\mathbf{ab}) = \pi(\mathbf{a})\pi(\mathbf{b})$ and $\pi(\mathbf{a}^*) = \pi(\mathbf{a})^*$. Unfortunately, for spin chains there is not a unique representation that can be used for all purposes, but one has to choose the representation, which is most appropriate to the given physical context. This ambiguity also reflects on the definition of states. In fact, one might be tempted to use density operators ρ on the Hilbert space $\mathbb{H}$, which carry the representation π. However, since different representations correspond to different physical contexts one should use all possible representations (in fact, each density operator in any representation can describe a state). Clearly, it would be much better to describe states in a way independent from the representation. Thus, a state of $\mathcal{A}_{\mathbb{Z}}$ is defined as a linear functional $\psi : \mathcal{A}_{\mathbb{Z}} \to \mathbb{C}$, which is positive ($\psi(\mathbf{a}^*\mathbf{a}) \geq 0$, for all $\mathbf{a} \in \mathcal{A}_{\mathbb{Z}}$) and normalized ($\psi(\mathbf{I}) = 1$). This means that given a representation ϕ and a density operator ρ on $\mathbb{H}$, the corresponding state is the functional $\psi_\rho(\mathbf{a}) = \mathrm{tr}[\pi(\mathbf{a})\rho]$. The possibility to find for each state ψ a Hilbert space $\mathbb{H}$ carrying a representation π and a density operator ρ such that $\psi = \psi_\rho$ is guaranteed by the Gelfand–Naimark– Segal theorem (see

Proposition 2.5.2). It states that each state ψ can be represented by a state vector $|v_\psi\rangle$ on a suitable Hilbert space. In other words, it is like to say that it is always possible to provide a 'purification' of the state ψ.

On the quasilocal algebra $\mathcal{A}_\mathbb{Z}$, we introduce a *shift operator* $\sigma : \mathcal{A}_\mathbb{Z} \to \mathcal{A}_\mathbb{Z}$ by setting

$$\sigma(\mathbf{a}) = \mathbf{I}_\mathcal{A} \otimes \mathbf{a} \in \mathcal{A}_{\mathbb{A}+1}, \quad \forall \mathbf{a} \in \mathcal{A}_\mathbb{A}, \tag{14.26}$$

where we have used the notation $\mathbb{A} + 1 := \{z + 1 \mid z \in \mathbb{A}\}$ and $\mathbf{A} \otimes \mathbf{I}_\mathcal{A}$ (resp., $\mathbf{I}_\mathcal{A} \otimes \mathbf{A}$) stands for the tensor product between $\mathbf{A}$ belonging to $\mathcal{A}_\mathbb{A}$ and identity of $\mathcal{A}$ on the site to the right of $\mathbb{A}$ (resp., between identity of $\mathcal{A}$ on the site to the left of $\mathbb{A}$ and a belonging to $\mathcal{A}_\mathbb{A}$). Moving from the action of the shift σ, it is possible to introduce the notion of the stationary state ψ on $\mathcal{A}_\mathbb{Z}$ when $\psi \circ \sigma = \psi$ holds true. The set of stationary states on $\mathcal{A}_\mathbb{Z}$ turns out to be convex. Then a state ψ on $\mathcal{A}_\mathbb{Z}$ is called ergodic (with respect to the shift) if it is extremal on this set.

The canonical extension of σ onto the quasilocal algebra $\mathcal{A}_\mathbb{Z}$ is a *-automorphism on $\mathcal{A}_\mathbb{Z}$, and the integer powers $\{\sigma^z, z \in \mathbb{Z}\}$ represent an action of the *translation group* $\mathbb{Z}$ by automorphisms on $\mathcal{A}_\mathbb{Z}$.

A quantum state ω on the infinite chain is a positive and normalized linear functional on $\mathcal{A}_\mathbb{Z}$. Equivalently, a quantum state ω is given by a family $\{\omega_\mathbb{A}, \mathbb{A} \subset \mathbb{Z}\}$ of density operators on $\mathcal{A}_\mathbb{A}$ for finite $\mathbb{A} \subset \mathbb{Z}$ such that $\omega(\mathbf{a}) = \mathrm{tr}[\omega_\mathbb{A}\mathbf{a}]$ for $\mathbf{a} \in \mathcal{A}_\mathbb{A}$. The local density operators have to satisfy the consistency condition that $\mathrm{tr}_{\mathbb{A}_2\setminus\mathbb{A}_1}[\omega_{\mathbb{A}_2}] = \omega_{\mathbb{A}_1}$ whenever $\mathbb{A}_1 \subset \mathbb{A}_2$. This equivalence reflects the fact that the quantum state of the entire infinite chain is assumed to be determined by the expectation values of all observables on finite subsystems $\mathbb{A} \subset \mathbb{Z}$.

14.3.2 Structure of causal channels

To set the stage for treating memory effects of quantum communication, we assume that Alice, the sender, has at her disposal a quantum channel to transmit at every discrete time step n a sequence of time-ordered classical messages $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{N_n}\}$, using block coding of length n and size N_n, to the receiver Bob. Recall that in the Schrodinger picture, a quantum channel is a completely positive, trace preserving linear map that transforms input states in $\mathcal{S}(\mathbb{H}_A)$ to output states in $\mathcal{S}(\mathbb{H}_B)$. In the quasilocal algebra approach for memory effects of quantum communication, we consider its corresponding dual channel in Heisenberg's picture. The dual channel is a completely positive and unital map that transforms quantum observables from observable algebra at receiver's end B to quantum observables from observable algebra at the sender's send A (see Definition 5.2.1).

In the following, we explore the quasilocal approach constructed in Subsection 14.3.1 to the quantum communication with memory. In this content, the dual channel is represented (in the Heisenberg picture) by a completely positive and unital map $\Upsilon : \mathcal{B}_{\mathbb{Z}} \to \mathcal{A}_{\mathbb{Z}}$ between the quasilocal algebras $\mathcal{B}_{\mathbb{Z}}$ and $\mathcal{A}_{\mathbb{Z}}$ on Bob's and Alice's side of the channel, respectively. In the following, we will restrict ourselves to translational invariant channels, i. e., we assume that Υ commutes with the shift on the spin chain: $\sigma_A \circ \Upsilon = \Upsilon \circ \sigma_B$. In addition, we impose the physically reasonable constraint that outputs up to some time t do not depend on inputs at times $t' > t$, leading to the following definition.

Definition 14.3.1 (Causal channels). A causal channel $\Upsilon : \mathcal{B}_{\mathbb{Z}} \to \mathcal{A}_{\mathbb{Z}}$ is a completely positive and unital translational invariant map such that for every $z \in \mathbb{Z}$,

$$\Upsilon(B_{]-\infty,z]} \otimes \mathbf{1}_{[z+1,+\infty[}) = \Upsilon(B_{]-\infty,z]}) \otimes \mathbf{1}_{[z+1,+\infty[} \tag{14.27}$$

for all $B_{]-\infty,z]} \in \mathfrak{B}_{]-\infty,z]}$, where $\mathfrak{B}_{]-\infty,z]}$ denotes the set of bounded operators defined on lattice elements up to that associated with the label z.

Note that memoryless configurations are obtained if in addition to (14.27) the following condition:

$$\Upsilon(\mathbf{1}_{]-\infty,z]} \otimes B_{[z+1,\infty[}) = \mathbf{1}_{]-\infty,z]} \otimes \Upsilon(B_{[z+1,\infty[}) \tag{14.28}$$

also holds for all $B_{[z+1,\infty[} \in \mathfrak{B}_{[1+z,\infty[}$.

Since Υ is translational invariant, we can henceforth set $z = 0$, and we will use the shorthand $\mathcal{A}_- := \mathcal{A}_{]-\infty,0]}$ and $\mathcal{A}_+ := \mathcal{A}_{[1,+\infty[}$ to denote the left- and right-half chain, respectively. $\mathfrak{B}_-$ and $\mathfrak{B}_+$ are defined analogously as the set of bounded operators on $]-\infty, 0]$ and $[1, \infty[$, respectively.

It can be shown that a concatenated memory channel is a causal channel.

The following structure theorem due originally to Kretschmann and Werner [100] states that every causal channel can be represented as a concatenated memory channel.

Theorem 14.3.2 (Structure theorem). *Let $\Upsilon : \mathcal{B}_{\mathbb{Z}} \to \mathcal{A}_{\mathbb{Z}}$ be a causal channel. Then, by ignoring its outputs on the left-hand chain $\mathcal{B}_-$, there exist a memory observable algebra $\mathcal{M}$ and an initializing channel $\mathfrak{R} : \mathcal{M} \to \mathcal{A}_-$ such that for all $n \in \mathbb{N}$,*

$$\Upsilon(\mathbf{1}_- \otimes \mathbf{B}_n) = (\mathfrak{R} \otimes \mathfrak{J}_{\mathcal{A}}^{\otimes n})\Phi^{*(n)}(\mathbf{B}_n \otimes \mathbf{1}_{\mathcal{M}}) \tag{14.29}$$

for all $\mathbf{B}_n \in \mathfrak{B}_{[1,n]} \cong \mathfrak{B}^{\otimes n}$, where $\Phi^{*(n)}$ is the n-fold concatenation of a memory channel $\Phi^* : \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ as defined in (14.17).

Proof. Let $\mathbb{H}$ be the Hilbert space associated with the universal representation of the left-half chain $\mathcal{A}_-$. Note that in general $\mathbb{H}$ will not be separable. However, separability is not required in Stinespring's theorem. Suppose that $(\mathbb{K}, \pi, \mathbf{V})$ is a minimal Stinespring dilation for $\Upsilon|_{\mathcal{B}_-}$, i. e.,

$$\Upsilon(b) = \mathbf{V}^*\pi(b)\mathbf{V}, \quad \forall b \in \mathcal{B}_- \tag{14.30}$$

for some Stinespring isometry $\mathbf{V} : \mathbb{H} \to \mathbb{K}$, where $\mathbb{K}$ is a certain Hilbert space, $\pi : \mathcal{B}_- \to \mathbb{K}$ is a *-homomorphism and $\mathbf{V}^* : \mathbb{K} \to \mathbb{H}$ is the adjoint operator of $\mathbf{V}$. From Stinespring's representation (14.30) and the causality property (14.27), we conclude that

$$\begin{aligned}\mathbf{V}^*\pi(\mathbf{b} \otimes \mathbf{1}_{\mathcal{B}}^{\otimes n})\mathbf{V} &= \Upsilon(\mathbf{b} \otimes \mathbf{1}_{\mathcal{B}}^{\otimes n}) = \Upsilon(\mathbf{b}) \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n} \\ &= (\mathbf{V}^*\pi(b)\mathbf{V}) \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n} = (\mathbf{V}^* \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n})(\pi(\mathbf{b}) \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n})(\mathbf{V} \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n}),\end{aligned} \tag{14.31}$$

for all $\mathbf{b} \in \mathcal{B}_-$. Since $\mathbf{V}$ is a minimal dilation for Υ so is $\mathbf{V} \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n}$ for $\Upsilon \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n}$. As explained earlier, we conclude that there exists an isometry $\mathbf{W}_n : \mathbb{K} \otimes \mathbb{C}_d^{\otimes n} \to \mathbb{K}$ defined by

$$\mathbf{W}_n(\pi(\mathbf{b}) \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n})(\mathbf{V} \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n})(\psi \otimes \psi_n) = \pi(\mathbf{b} \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n})\mathbf{V}(\psi \otimes \psi_n) \tag{14.32}$$

for all $\mathbf{b} \in \mathcal{B}_-$, $\psi \in \mathbb{H}$, and $\psi_n \in \mathcal{A}^{\otimes n}$ such that

$$\pi(\mathbf{b} \otimes \mathbf{1}_{\mathcal{B}}^{\otimes n})\mathbf{W}_n = \mathbf{W}_n(\pi(\mathbf{b}) \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n}) \tag{14.33}$$

for all $\mathbf{b} \in \mathcal{B}_-$, and

$$\mathbf{W}_n(\mathbf{V} \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n}) = \mathbf{V}. \tag{14.34}$$

We now reconstruct the memory algebra as follows. Let $\mathcal{M} := \pi'(\mathcal{B}_-)$, the *commutant* of the observable algebra $\mathcal{B}_-$, and let $\mathfrak{S}_n : \mathcal{B}^{\otimes n} \otimes \mathcal{M} \to \mathfrak{B}(\mathbb{K}) \otimes \mathfrak{B}(\mathbb{C}_d^{\otimes n})$ be defined by

$$\Phi^{*(n)}(\mathbf{b} \otimes \mathbf{m}) := \mathbf{W}_n^*\pi(\mathbf{b})\mathbf{m}\mathbf{W}_n \tag{14.35}$$

for all $\mathbf{b} \in \mathcal{B}_-$ and $\mathbf{m} \in \mathcal{M}$. The memory initializing channel $\mathfrak{R} : \mathcal{M} \to \mathcal{A}_-$ is given by

$$\mathfrak{R}(\mathbf{m}) := \mathbf{V}^*\mathbf{m}\mathbf{V}, \quad \forall \mathbf{m} \in \mathcal{M}. \tag{14.36}$$

In order to justify these choices, we will first show that

$$\Phi^{*(n)}(\mathcal{B}^{\otimes n} \otimes \mathcal{M}) \subset \mathcal{M} \otimes \mathcal{A}^{\otimes n}. \tag{14.37}$$

Noting that $\pi(\mathbf{1}_{\mathcal{B}_-} \otimes \mathcal{B}^{\otimes n})\mathcal{M} \subset \pi'(\mathcal{B}_- \otimes \mathbf{1}_{\mathcal{B}}^{\otimes n})$, we see from (14.33) that

$$\begin{aligned}&\mathbf{W}_n^*\pi(\mathbf{1}_{\mathcal{B}_-} \otimes \mathbf{b}_n)\mathbf{m}\mathbf{W}_n(\pi(\tilde{\mathbf{b}}_{\mathcal{B}_-}) \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n}) \\ &\quad = \mathbf{W}_n^*\pi(\mathbf{1}_{\mathcal{B}_-} \otimes \mathbf{b}_n)\mathbf{m}\pi(\tilde{\mathbf{b}}_{\mathcal{B}_-} \otimes \mathbf{1}_{\mathcal{B}}^{\otimes n})\mathbf{W}_n \\ &\quad = \mathbf{W}_n^*\pi(\tilde{\mathbf{b}}_{\mathcal{B}_-} \otimes \mathbf{1}_{\mathcal{B}}^{\otimes n})\pi(\mathbf{1}_{\mathcal{B}_-} \otimes \mathbf{b}_n)\mathbf{m}\mathbf{W}_n \\ &\quad = (\pi(\tilde{\mathbf{b}}_{\mathcal{B}_-}) \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n})\mathbf{W}_n^*\pi(\mathbf{1}_{\mathcal{B}_-} \otimes \mathbf{b}_n)\mathbf{m}\mathbf{W}_n\end{aligned} \tag{14.38}$$

for all $\mathbf{b}_n \in \mathcal{B}^{\otimes n}$ and $\tilde{\mathbf{b}}_{\mathcal{B}_-} \in \mathcal{B}_-$, implying that

$$[\Phi^{*(n)}(\mathbf{b} \otimes \mathbf{m}) \mid \pi(\tilde{\mathbf{b}}_{\mathcal{B}_-}) \otimes \mathbf{1}_{\mathcal{A}}^{\otimes n}] = 0, \tag{14.39}$$

from which (14.37) follows. To complete the proof, it suffices to show that $\Phi^{(n)}$ has the right concatenation properties, i. e.,

$$\mathfrak{R}(\mathbf{m}) = (\mathfrak{R} \otimes \mathbf{I}_{\mathcal{A}}^{\otimes n})\Phi^{*(n)}(\mathbf{1}_{\mathcal{B}}^{\otimes n} \otimes \mathbf{m}) \tag{14.40}$$

and

$$\Upsilon(\mathbf{b}) = (\mathfrak{R} \otimes \mathbf{I}_{\mathcal{A}}^{\otimes n})\Phi^{*(n)}(\mathbf{b} \otimes \mathbf{1}_{\mathcal{M}}) \tag{14.41}$$

for all $\mathbf{m} \in \mathcal{M}$ and $\mathbf{b} \in \mathcal{B}^{\otimes n}$. However, this is immediately from the definitions of $\Phi^{(n)}$ and $\mathfrak{R}$ and (14.35). The result then follows by setting $\Phi^{*(1)} := \Phi^*$. This proves the theorem. □

14.4 Forgetful channels

In this section, we explore a special but important class of quantum memory channels, namely, the class of *forgetful channels*. Forgetful channels have been studied by Bowen and Mancini [14] and more recently by Kretschmann and Werner [100].

The presentation of the concept and properties of forgetful channels in this section can be found in Kretschmann and Werner [100] (see also Caruso et al. [18]).

14.4.1 Definition of forgetful channels

Forgetful channels have been studied by Bowen and Mancini [14] and more recently by Kretschmann and Werner [100].

Consider a quantum memory channel in the Heisenberg picture $\Phi^* : \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$, where $\mathcal{A} = \mathfrak{B}(\mathbb{H}_A)$, $\mathcal{B} = \mathfrak{B}(\mathbb{H}_B)$ and $\mathcal{M} = \mathfrak{B}(\mathbb{H}_M)$, respectively, represent C^*-algebras of operators on systems A, B and the memory system M. The n-concatenation of the channel Φ^* is the channel $\Phi^{*(n)} : \mathcal{B}^{\otimes n} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}^{\otimes n}$ given as

$$\Phi^{*(n)} = (\Phi^* \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes(n-1)}) \circ \cdots \circ (\mathfrak{I}_{\mathcal{B}}^{\otimes(n-2)} \otimes \Phi^* \otimes \mathfrak{I}_{\mathcal{A}}) \circ (\mathfrak{I}_{\mathcal{B}}^{\otimes(n-1)} \otimes \Phi^*),$$

where $\mathfrak{I}_{\mathcal{A}}$ and $\mathfrak{I}_{\mathcal{B}}$ denote the identity operator on $\mathcal{A} = \mathfrak{B}(\mathbb{H}_A)$ and $\mathcal{B} = \mathfrak{B}(\mathbb{H}_B)$, respectively.

Forgetful channels in the Heisenberg picture are quantum memory channels $\Phi^* : \mathfrak{B}(\mathbb{H}_B) \otimes \mathcal{M} \to \mathcal{M} \otimes \mathfrak{B}(\mathbb{H}_A)$ in which the effect of the initializing memory state dies away with time. More formally, Kretschmann and Werner [100] give the following definition.

Definition 14.4.1. Let $\Phi^* : \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ be a (dual) quantum memory channel, where $\mathcal{A} = \mathfrak{B}(\mathbb{H}_A)$ and $\mathcal{B} = \mathfrak{B}(\mathbb{H}_B)$, and let $\Phi^{*(n)} : \mathcal{B}^{\otimes n} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}^{\otimes n}$ be its n-fold concatenation. Let $\hat{\Phi}^{*(n)} : \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}^{\otimes n}$ be the concatenated channel in which Bob's output is ignored: $\hat{\Phi}^{*(n)}(\mathfrak{m}) := \Phi^{*(n)}(\mathfrak{I}_{\mathcal{B}} \otimes \mathfrak{m})$ for all $\mathfrak{m} \in \mathcal{M}$, where $\mathfrak{I}_{\mathcal{B}}$ is the identity operator on $\mathcal{B}$. The quantum memory channel Φ^* is said to be *forgetful* if there exists a sequence of channels $\tilde{\Phi}_n^* : \mathcal{M} \to \mathcal{A}^{\otimes n}$ such that

$$\lim_{n \to +\infty} \|\hat{\Phi}^{*(n)} - \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_n^*\|_{cb} = 0, \tag{14.42}$$

where $\| \cdot \|_{cb}$ denotes the complete bounded norm defined in Definition 6.3.1 and $\mathfrak{I}_{\mathcal{M}}$ is the identity operator on $\mathcal{M}$.

In other words, a quantum channel Φ^* (or equivalently Φ) is called forgetful if the memory behavior does not depend on the initial memory configuration. That is, if for any input state $\rho^{(n)}$ and $\epsilon > 0$ there exists an $N(\epsilon)$ such that for all $n \geq N(\epsilon)$,

$$\left\|\Phi^{*(n)}(\rho^{(n)} \otimes \omega) - \Phi^{*(n)}(\rho^{(n)} \otimes \sigma)\right\|_1 < \epsilon \tag{14.43}$$

for any pair of initial memory ω and σ, in which $\Phi^{*(n)} : \mathcal{B}^{\otimes n} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}^{\otimes n}$ denotes concatenated channel given as

$$\Phi^{*(n)} = (\Phi^* \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes(n-1)}) \circ \cdots \circ (\mathfrak{I}_{\mathcal{B}}^{\otimes(n-2)} \otimes \Phi^* \otimes \mathfrak{I}_{\mathcal{A}}) \circ (\mathfrak{I}_{\mathcal{B}}^{\otimes(n-1)} \otimes \Phi^*),$$

where $\mathfrak{I}_{\mathcal{A}}$ and $\mathfrak{I}_{\mathcal{B}}$ denote the identity operator on $\mathcal{A} = \mathfrak{B}(\mathbb{H}_A)$ and $\mathcal{B} = \mathfrak{B}(\mathbb{H}_B)$, respectively.

The following example can be found in Kretschmann and Werner [100].

Example 14.2. Consider the classically mixed channel $\Phi = p\mathfrak{I} + (1-p)\sigma$, where $p \in [0,1[$ and σ is the shift operator introduced in (14.26). When this channel is concatenated, in every step either the ideal channel $\mathfrak{I}$ or the shift channel σ is chosen with probabilities p and $1-p$, respectively. The only possible way for an n-fold concatenation $\hat{\Phi}^{*(n)}$ not to be forgetful is to choose the ideal channel $\mathfrak{I}$ in every step. However, the probability for this event is p^n, and thus vanishes in the limit $n \to \infty$, implying that (14.42) holds.

Example 14.3. A forgetful channel is a special case of the quantum channel with Markovian memory. Specifically, if the Markov transition matrix $\{q_{ij}\}_{i,j\in\mathbb{I}}$ is stationary, the quantum channel with Markovian memory Φ_∞ defined in (15.4) is a forgetful channel.

Remark 14.1. Note that Definition 14.4.1 for the forgetful channel can be relaxed by requiring that $(\tilde{\Phi}_n)_{n=1}^\infty$ being a sequence of linear maps only. To see that this leads to the equivalence of the definition of forgetfulness, assume that $\|\hat{\Phi}^{*(n)} - \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_n^*\|_{cb} < \epsilon$ for some $\epsilon > 0$ and $n \geq 0$ and some linear operator $\tilde{\Phi}_n$. Replacing $\mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_n^*$ with the quantum channel $(\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes n})\circ\hat{\Phi}^{*(n)}$, where $\mathfrak{P}:\mathcal{M}\to\mathbb{C}\mathfrak{I}_{\mathcal{M}}$ is the completely dephasing channel (see Wilde [178] and Watrous [173] for a definition and examples of dephasing channels), we see that

$$\begin{aligned}
&\|\hat{\Phi}^{*(n)}-(\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes n})\circ\hat{\Phi}^{*(n)}\|_{cb}\\
&\le\|\hat{\Phi}^{*(n)}-\mathfrak{I}_{\mathcal{M}}\otimes\tilde{\Phi}_n^*\|_{cb}+\|(\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes n})\circ(\mathfrak{I}_{\mathcal{M}}\otimes\tilde{\Phi}_n^*-\hat{\Phi}^{*(n)})\|_{cb}\\
&\le 2\|\hat{\Phi}^{*(n)}-\mathfrak{I}_{\mathcal{M}}\otimes\tilde{\Phi}_n^*\|_{cb}\le 2\epsilon,
\end{aligned}\tag{14.44}$$

where thus $\lim_{n\to+\infty}\|\hat{\Phi}^{*(n)}-(\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes n})\circ\hat{\Phi}^{*(n)}\|_{cb}=0$, implying the channel Φ is indeed forgetful by Definition 14.4.1.

The following lemmas are needed for proving Proposition 14.4.4, which characterizes a forgetful channel.

Lemma 14.4.2. *Let $(d_n)_{n=1}^{\infty}$ be a positive and nonincreasing sequence satisfying the subadditivity inequality*

$$d_{n+m}\le d_n d_m,\quad \forall n,m\in\mathbb{N}.\tag{14.45}$$

Assume further that $d_N<1$ for some $N\in\mathbb{N}$. Then

$$d_n\le c^n,\quad \forall n\ge N\tag{14.46}$$

for some constant $c<1$.

Proof. Assume that $d_N<1$ for some $N\in\mathbb{N}$. From the subadditivity inequality (14.45), we then see that $d_{N+N}\le d_N^2$, and by induction, $d_{\nu N}\le d_N^{\nu}$ for all $\nu\in\mathbb{N}$. By the monotonicity of $(d_n)_{n=1}^{\infty}$, we may then conclude that for $n\in[\nu N,(\nu+1)N[$ we have

$$d_n\le d_{\nu N}\le d_N^{\nu}\le\big(d_N^{\frac{1}{2N}}\big)^n=c^n\tag{14.47}$$

with $c:=d_N^{\frac{1}{2N}}<1$. This proves the lemma. □

Lemma 14.4.3. *Let $\mathfrak{R}:\mathfrak{B}(\mathbb{H}_M)\to\mathcal{A}$ be a linear operator, and assume that $d_M:=\dim(\mathbb{H}_M)<\infty$. We then have*

$$\|\mathfrak{R}\|_{cb}\le d_M^2\|\mathfrak{R}\|_{\infty}.\tag{14.48}$$

Proof. By definition of the cb-norm $\|\cdot\|_{cb}$ (see Definition 6.3.1), we have $\|\mathfrak{R}\|_{cb}=\sup_k\{\|\mathfrak{R}\otimes\mathfrak{I}_k\|_{\infty}\}$, where $\mathfrak{I}_k$ is the identity operator of $k\times k$ matrices $\mathfrak{B}(\mathbb{C}_k)$. Every $\mathbf{x}\in\mathfrak{B}(\mathbb{H}_M)\otimes\mathfrak{B}(\mathbb{C}_k)$ can have the expansion of the form

$$\mathbf{x}=\sum_{\alpha}\mathbf{m}_{\alpha}\otimes\mathbf{k}_{\alpha}=\sum_{\alpha}\sum_{i,j=1}^{d_M}\mu_{\alpha,ij}|i\rangle_M\langle j|\otimes\mathbf{k}_{\alpha}=\sum_{i,j=1}^{d_M}|i\rangle_M\langle j|\mathbf{x}_{ij},\tag{14.49}$$

where $\{|i\rangle_M\}_{i=1}^{d_M}$ is an orthonormal basis for $\mathbb{H}_M$, and $\mathbf{x}_{ij} := \sum_\alpha \mu_{\alpha,ij}\mathbf{k}_\alpha$. Note that $\|\mathbf{x}_{ij}\|_\infty \leq \|\mathbf{x}\|_\infty$ for all $i, j = 1, 2, \ldots, d_M$, implying that

$$\begin{aligned}\|(\mathfrak{R} \otimes \mathfrak{I}_k)(\mathbf{x})\|_\infty &= \left\| \sum_{i,j=1}^{d_M} \mathfrak{R}(|i\rangle_M\langle j|) \otimes \mathbf{x}_{ij} \right\|_\infty \\ &\leq \sum_{i,j=1}^{d_M} \|\mathfrak{R}\|_\infty \||i\rangle_M\langle j|\|_\infty \|\mathbf{x}_{ij}\|_\infty \leq d_M^2 \|\mathfrak{R}\|_\infty \|\mathbf{x}\|_\infty \end{aligned} \tag{14.50}$$

holds independent of k. Consequently, we have

$$\|\mathfrak{R}\|_{cb} = \sup_k\{\|\mathfrak{R} \otimes \mathfrak{I}_k\|_\infty\} \leq d_M^2 \|\mathfrak{R}\|_\infty.$$

This proves the lemma. □

There exist several equivalent criteria for a quantum memory channel to be forgetful. In particular, to show the channel Φ is forgetful, it is sufficient to show that the norm distance $\|\hat{\Phi}^{*(n)} - \mathfrak{I}_\mathcal{M} \otimes \tilde{\Phi}_n\|_{cb} < 1$ for some $n \in \mathbb{N}$. What is more important is that the memory effects can always be assumed to vanish exponentially fast. In addition, if the memory algebra $\mathcal{M}$ has finite dimension, the cb-norm criterion (14.42) can be replaced by the usual operator norm $\|\cdot\|_\infty$. In fact, we have the following result.

Proposition 14.4.4. *Let* $\Phi^* : \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ *be a quantum memory channel, and for* $n \in \mathbb{N}$, *let* $\hat{\Phi}^{*(n)}$ *be as defined in Definition* 14.4.1. *Then the quantum memory channel* Φ^* *is forgetful if and only if there exists an integer* $N \in \mathbb{N}$ *and some linear operator* $\tilde{\Phi}_n : \mathcal{M} \to \mathcal{A}^{\otimes N}$ *(not necessary a channel) such that*

$$\left\|\hat{\Phi}^{*(N)} - \mathfrak{I}_\mathcal{M} \otimes \tilde{\Phi}_N\right\|_{cb} < 1. \tag{14.51}$$

Assume in addition that the memory algebra $\mathcal{M}$ *has finite dimension. Then* Φ^* *is forgetful if and only if for every* $\mathbf{m} \in \mathcal{M}$ *and* $\epsilon > 0$ *there exist a positive integer* $N \in \mathbb{N}$ *and* $\mathfrak{a}_N \in \mathcal{A}^{\otimes N}$ *such that*

$$\left\|\hat{\Phi}^{*(N)} - \mathfrak{I}_\mathcal{M} \otimes \mathfrak{a}_N\right\|_\infty \leq \epsilon \|\mathfrak{m}\|_\infty. \tag{14.52}$$

Proof. 1. We first prove part one of the proposition, where no assumption on the dimensionality of $\mathcal{M}$ is made. If Φ is forgetful, (14.51) is immediate from the definition. In order to prove the converse, let

$$d_n := \inf\{\left\|\hat{\Phi}^{*(n)} - \mathfrak{I}_\mathcal{M}\tilde{\Phi}_n\right\|_{cb} \mid \tilde{\Phi}_n : \mathcal{M} \to \mathcal{A}^{\otimes n} \text{ is a linear map}\}, \tag{14.53}$$

for $n \in \mathbb{N}$. Our strategy is to show that the sequence $(d_n)_{n=1}^{+\infty}$ satisfies the conditions of Lemma 14.4.2. From (14.51), we can then conclude that $d_n \leq c^n$ for all $n \geq N$ for some constant $c < 1$, and thus Φ is forgetful with exponentially vanishing errors by Remark 15.1.

We start by showing that the sequence $(d_n)_{n=1}^{+\infty}$ is nonincreasing, i. e., $d_{n+1} \le d_n$ for all $n \in \mathbb{N}$. From the definition of $\hat{\Phi}^{*(n)}$, we have

$$\begin{aligned}\hat{\Phi}^{*(n+1)} &= (\hat{\Phi}^* \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes n}) \circ \hat{\Phi}^{*(n)} \\ &= (\hat{\Phi}^* \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes n}) \circ (\hat{\Phi}^{*(n)} - \mathfrak{I}_{\mathcal{M}}\tilde{\Phi}_n) + (\hat{\Phi}^* \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes n}) \circ (\mathfrak{I}_{\mathcal{M}}\tilde{\Phi}_n) \\ &= (\hat{\Phi}^* \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes n})(\hat{\Phi}^{*(n)} - \mathfrak{I}_{\mathcal{M}}\tilde{\Phi}_n) + \mathfrak{I}_{\mathcal{M}} \otimes \mathfrak{I}_{\mathcal{A}} \otimes \tilde{\Phi}_n,\end{aligned} \tag{14.54}$$

where in the last step we have applied the unitarity of the channel $\hat{\Phi}$. From (14.53) and unitality of the cb-norm, we may conclude that

$$\begin{aligned}d_{n+1} &\le \|\hat{\Phi}^{*(n+1)} - \mathfrak{I}_{\mathcal{M}} \otimes \mathfrak{I}_{\mathcal{A}} \otimes \tilde{\Phi}_n\|_{cb} \\ &\le \|\hat{\Phi}^* \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes n}\|_{cb}\|\hat{\Phi}^{*(n)} - \mathfrak{I}_{\mathcal{M}}\tilde{\Phi}_n\|_{cb} \le d_n.\end{aligned}$$

This proves $d_{n+1} \le d_n$. We now want to prove $d_{n+m} \le d_m d_n$ for all $n, m \in \mathbb{N}$. Similar to the above estimate, we have

$$\begin{aligned}\hat{\Phi}^{*(n+m)} &= (\hat{\Phi}^{*(n)} \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes m}) \circ \hat{\Phi}^{*(m)} \\ &= (\hat{\Phi}^{*(n)} \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes m})(\hat{\Phi}^{*(m)} - \mathfrak{I}_{\mathcal{M}}\tilde{\Phi}_m) + (\hat{\Phi}^{*(n)} \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes m}) \circ (\mathfrak{I}_{\mathcal{M}}\tilde{\Phi}_m) \\ &= [(\hat{\Phi}^{*(n)} - \mathfrak{I}_{\mathcal{M}}\tilde{\Phi}_n) \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes m}](\hat{\Phi}^{*(m)} - \mathfrak{I}_{\mathcal{M}}\tilde{\Phi}_m) + \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_{n+m},\end{aligned} \tag{14.55}$$

where we have introduced the shorthand

$$\tilde{\Phi}_{n+m} := \mathfrak{I}_{\mathcal{A}}^{\otimes n} \otimes \tilde{\Phi}_m + (\tilde{\Phi}_n \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes m})(\hat{\Phi}^{*(m)} - \mathfrak{I}_{\mathcal{M}}^{\otimes m}\tilde{\Phi}_m). \tag{14.56}$$

Invoking again the unitality and multiplicativity of the cb-norm $\|\cdot\|_{cb}$, we conclude from (14.55) that

$$\begin{aligned}&\|\hat{\Phi}^{*(n+m)} - \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_{n+m}\|_{cb} \\ &\quad \le \|\hat{\Phi}^{*(n)} - \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_n\|_{cb}\|\hat{\Phi}^{*(m)} - \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_m\|_{cb} \le d_n d_m.\end{aligned} \tag{14.57}$$

This proves the estimate. Note that $\tilde{\Phi}_{n+m}$ is clearly a linear and unital but not necessarily positive. This is why we did not require the maps $\Phi^{*(n)}$ to be channels in the definition of the sequence $(d_n)_{n=1}^{\infty}$. This completes the first part of the proof.

2. For the second part, assume that $\mathcal{M} = \mathfrak{B}(\mathbb{H}_M)$ with $d_M := \dim(\mathbb{H}_M) < \infty$. If (14.52) holds, by the same reasoning as in Remark 15.1, we may conclude that $\mathfrak{I}_{\mathcal{M}} \otimes \mathfrak{a}_N$ can be replaced by $(\mathfrak{P} \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes N}) \circ \hat{\Phi}^{*(N)}(\mathbf{m})$, implying that for every $\epsilon > 0$ and every $\mathbf{m} \in \mathcal{M}$, we can find a positive integer $N \in \mathbb{N}$ such that

$$\|\hat{\Phi}^{*(N)}(\mathbf{m}) - (\mathfrak{P} \otimes \mathfrak{I}_{\mathcal{A}}^{\otimes N})(\mathbf{m})\|_{\infty} \le 2\epsilon\|\mathbf{m}\|_{\infty}. \tag{14.58}$$

In order to arrive at the uniform bound, we introduce an orthonormal basis $\{|i\rangle\}_{i=1}^{d_M}$ for $\mathbb{H}_M$. Since $\mathbb{H}_M$ is finite-dimensional, (14.58) holds uniformly for the basis operators $\{|i\rangle_M\langle j|\}_{i,j=1}^{d_M}$ for some possibly larger N. Thus, by setting $\mathbf{m} = \sum_{i,j=1}^{d_M} m_{ij}|i\rangle_M\langle j|$, we see that

$$\begin{aligned}
&\|\hat{\Phi}^{*(N)}(\mathbf{m}) - (\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes N})\circ\hat{\Phi}^{*(N)}(\mathbf{m})\|_\infty \\
&\leq \sum_{i,j=1}^{d_M} |m_{ij}|\|\hat{\Phi}^{*(N)}(|i\rangle_M\langle j|) - (\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes N})\circ\hat{\Phi}^{*(N)}(|i\rangle_M\langle j|)\|_\infty \\
&\leq 2\epsilon\sum_{i,j=1}^{d_M} |m_{ij} \leq 2\epsilon d_M^2\|\mathbf{m}\|_\infty,
\end{aligned} \tag{14.59}$$

where in the last step we have used that $|m_{ij}| \leq \|\mathbf{m}\|_\infty$ for all $i,j = 1,\dots,d_M$. Making use of Lemma 14.4.3, we may conclude from (14.59) that

$$\|\hat{\Phi}^{*(N)} - (\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes N})\circ\hat{\Phi}^{*(N)}\|_{cb} \leq 2\epsilon d_M^4. \tag{14.60}$$

Thus, choosing $\epsilon < \frac{1}{2d_M^4}$, we may choose $N \in \mathbb{N}$ such that (14.51) holds. Therefore, Φ is forgetful by the first part of the proof. The converse is immediate from the definition of forgetfulness. This proves the proposition. □

From the proof of the above proposition, we may immediately deduce the following.

Corollary 14.4.5. *Let $\Phi^* : \mathcal{B}\otimes\mathcal{M} \to \mathcal{M}\otimes\mathcal{A}$ be a forgetful quantum channel. Then the effect of the initial memory vanishes exponentially fast, i. e., we may find a constant $c < 1$ such that*

$$\|\hat{\Phi}^{*(n)} - (\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes n})\circ\hat{\Phi}^{*(n)}\|_{cb} < c^n \tag{14.61}$$

for all sufficiently large n.

Proposition 14.4.6. *Let $\Phi^* : \mathcal{B}\otimes\mathcal{M} \to \mathcal{M}\otimes\mathcal{A}$ be a quantum memory channel. Let $\epsilon > 0$, and for $n \in \mathbb{N}$, let $\hat{\Phi}^{*(n)}$ be defined as in Definition* 14.4.1. *Assume that*

$$\|\hat{\Phi}^{*(n)} - (\mathfrak{P}\otimes\mathfrak{I}_{\mathcal{A}}^{\otimes n})\circ\hat{\Phi}^{*(n)}\|_\infty \leq \epsilon, \tag{14.62}$$

where $\mathfrak{P} : \mathcal{M} \to \mathbb{C}\mathfrak{I}_{\mathcal{M}}$ is completely depolarizing channel. We then have

$$\|\mathrm{tr}_{\mathcal{B}^{\otimes n}}[\Phi^{(n)}(\rho_1 - \rho_2)]\|_1 \leq 2\epsilon \tag{14.63}$$

for all $\rho_1, \rho_2 \in \mathcal{M}^*\otimes\mathcal{A}^*\otimes n$ such that $\mathrm{tr}_{\mathcal{M}}[\rho_1] = \mathrm{tr}_{\mathcal{M}}[\rho_2]$. Conversely, suppose (14.63) *holds. Then* (14.62) *holds with the substitution $\epsilon \mapsto 2\epsilon$.*

In particular, if the quantum channel Φ is forgetful, then from Remark 16.1 we know that the condition in (14.62) is satisfied, and thus (14.63) holds. If in addition the memory algebra $\mathcal{M}$ is finite-dimensional, (14.62) is a necessary and sufficient criterion for forgetfulness by Proposition 14.4.4. By the above proposition, (14.63) then gives a necessary and sufficient criterion for forgetfulness in the Schrodinger picture language.

Proof. Note that for any linear operator $\mathfrak{T} : \mathcal{B} \to \mathcal{A}$, the operator norm $\|\mathfrak{T}\|_\infty$ equals the norm of the adjoint operator on the dual space, i. e.,

$$\|\mathfrak{T}\|_\infty = \sup_{\|\rho\|_1 \le 1} \left\|\mathfrak{T}^*(\rho)\right\|_1. \tag{14.64}$$

Suppose that (14.62) holds. Since $\mathfrak{I}_{\mathcal{A}}^{*\otimes n} \otimes \mathfrak{P}^*(\cdot) = \mathrm{tr}_{\mathcal{M}}[\cdot]$, the partial trace on the memory algebra $\mathcal{M}$, we may conclude from (14.62) and the norm duality (14.64) that

$$\hat{\Phi}^{(n)}(\rho) - \hat{\Phi}^{(n)}(\mathrm{tr}_{\mathcal{M}}[\rho])\|_1 \le \epsilon, \quad \forall \rho \in \mathcal{M}^* \otimes \mathcal{A}^{*\otimes n}, \tag{14.65}$$

which implies that for arbitrary $\rho_1, \rho_2 \in \mathcal{A}^* \otimes \mathcal{A}^{*\otimes n}$ such that $\mathrm{tr}_{\mathcal{M}}[\rho_1] = \mathrm{tr}_{\mathcal{M}}[\rho_2]$, we have

$$\left\|\hat{\Phi}^{(n)}(\rho_1) - \hat{\Phi}^{(n)}(\rho_2)\right\|_1 \le 2\epsilon \tag{14.66}$$

by application of triangle inequality. Equation (14.63) then follows by noting $\hat{\Phi}^{(n)} = \mathrm{tr}_{\mathcal{B}^{\otimes n}} \circ \hat{\Phi}^{(n)}$.

Conversely, from (14.63), we conclude that

$$\left\|\hat{\Phi}^{(n)}(\rho - \mathrm{tr}_{\mathcal{M}}[\rho])\right\|_1 \le 2\epsilon, \tag{14.67}$$

which implies (14.62) (with the substitution $\epsilon \to 2\epsilon$) by means of the norm duality (14.64). This proves the proposition. □

Proposition 14.4.4 and its Schrodinger dual proposition 14.4.6 can be employed to test whether a given quantum memory channel is forgetful as shown in the following example.

Example 14.4. Consider the unitary *partial flip operation*

$$\mathfrak{U}_\eta := (\cos\eta)\mathfrak{F} + i(\sin\eta)\mathfrak{I}, \quad i = \sqrt{-1}, \tag{14.68}$$

for $\eta \in [0, 2\pi[$, where $\mathfrak{F} : \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ denotes a *flip operation* defined by $\mathfrak{F} := \sum_{i,j} |ij\rangle\langle ji|$. Since $\mathfrak{F}(\mathbf{m} \otimes \mathbf{b})\mathfrak{F} = \mathbf{b} \otimes \mathbf{m}$ for all $\mathbf{b} \in \mathcal{B}$ and $\mathbf{m} \in \mathcal{M}$, $\mathfrak{U}_\eta = \mathfrak{F}$ when $\eta = 0$ is simply a shift operator, which has been proven to be a forgetful channel.

14.4.2 Topological properties

Let $\mathfrak{QC}_{\text{for}}(A, B)$ be the collection of forgetful channels and let $\mathfrak{QC}_{\text{mem}}(A, B)$ be the set of all quantum memory channels from system A to system B. In the following, we demonstrate that $\mathfrak{QC}_{\text{for}}(A, B)$ is a topologically important class of quantum memory channels by showing that $\mathfrak{QC}_{\text{for}}(A, B)$ is dense and open in $\mathfrak{QC}_{\text{mem}}(A, B)$ under $\|\cdot\|_{cb}$-norm. That is, for every nonforgetful quantum memory channel we may find a forgetful memory channel, which differs arbitrarily small from it under $\|\cdot\|_{cb}$-norm.

Theorem 14.4.7. *The set of forgetful quantum channels $\mathfrak{QC}_{\text{for}}(A, B)$ is open and dense in the set of quantum memory channels $\mathfrak{QC}_{\text{mem}}(A, B)$ in $\|\cdot\|_{cb}$-norm topology.*

Proof. 1. We first prove that $\mathfrak{QC}_{\text{for}}(A, B)$ is dense in $\mathfrak{QC}_{\text{mem}}(A, B)$ under the $\|\cdot\|_{cb}$-norm topology. Let $\Phi^* : \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$ be (in the Heisenberg picture) of any quantum memory channel. We can easily construct a forgetful channel by mixing it with the completely depolarizing channel

$$\mathfrak{D}(\mathbf{b} \otimes \mathbf{m}) := \operatorname{tr}[(\mathbf{b} \otimes \mathbf{m})\delta]\mathfrak{I}_{\mathfrak{M}} \otimes \mathcal{A}, \tag{14.69}$$

where $\delta \in \mathcal{B}^* \otimes \mathcal{M}^*$ is an arbitrary quantum state.

Just as in the classically mixed shift channel discussed above, all of the terms in an n-fold concatenation of the mixed channel $\Phi_\epsilon := (1 - \epsilon)\Phi + \epsilon\mathfrak{D}$ yield the identity operator $\mathcal{M}$ in the memory input, possibly apart from the $\Phi^{(n)}$-contribution, which scales as $(1 - \epsilon)n$, and thus vanishes as $n \to +\infty$. Since this holds for all $\epsilon > 0$, and $\|\Phi - \Phi_\epsilon\|_{cb} \leq 2\epsilon$, we have found a forgetful channel Φ_ϵ arbitrarily close to Φ, completing the proof of the first part of the theorem.

2. We will now show that the set of forgetful quantum channels $\mathfrak{QC}_{\text{for}}(A, B)$ is open. So, assume that we are given a forgetful memory channel $\Phi : \mathcal{B} \otimes \mathcal{M} \to \mathcal{M} \otimes \mathcal{A}$. We will show that Φ has a finite-size neighborhood in which all memory channels are forgetful. Clearly, by the definition of forgetfulness we can find $N \in \mathbb{N}$ and a quantum channel $\tilde{\Phi}_N : \mathcal{M} \to \mathcal{A}^{\otimes N}$ such that $\|\hat{\Phi}^{*(N)} - \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_N\|_{cb} < \frac{1}{2}$. Thus, for all quantum memory channels Ψ such that $\|\Psi^* - \Phi^*\|_{cb} \leq \frac{1}{2N}$, we have

$$\left\|\hat{\Psi}^{*(N)} - \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_N\right\|_{cb} \leq \left\|\hat{\Phi}^{*(N)} - \mathfrak{I}_{\mathcal{M}} \otimes \tilde{\Phi}_N\right\|_{cb} + N\left\|\Psi^* - \Phi^*\right\|_{cb} < 1 \tag{14.70}$$

and the forgetfulness of Φ^* immediately follows from Proposition 14.4.4. This proves the theorem. □

15 Channels with Markovian memory

This chapter constructs and explores classical capacities for channels with ergodic and general Markov memories.

15.1 A brief review on Markov chains

We first review some concepts of a homogeneous discrete time Markov chain with a finite state space from probability theory. The review material presented in this section can be found in the research monograph by Norris [117].

A discrete-time Markov chain is a sequence of random variables $\{X_n\}_{n=0}^{+\infty}$ defined on a probability space $(\Omega, \Sigma, \Pr)$ and taking its values in a state space $\mathbb{I}$ (which is assumed to be a finite set) with the following Markov property:

$$\begin{aligned}&\Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_1 = x_1, X_0 = x_0)\\&\quad = \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n),\end{aligned}$$

for all $n \geq 0$ and for all $x_i \in \mathbb{I}, i = 0, 1, 2, \dots, n$, if both conditional probabilities are well defined, i. e., if

$$\Pr(X_n = x_n, \dots, X_0 = x_0) > 0.$$

The Markov chain $\{X_n\}_{n=0}^{+\infty}$ is said to be time homogeneous if $P(X_{n+1} = x_{n+1}|X_n = x_n) = \cdots = P(X_1 = x_1|X_0 = x_0)$.

A. Probability transition matrix

The stochastic behavior such as the joint probability distribution $\Pr(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)$ of the Markov chain $\{X_n\}_{n=0}^{+\infty}$ is completely determined by the transition probability matrix $P = [q_{ij}]_{i,j\in\mathbb{I}}$ and its initial value X_0, where q_{ij} is the probability that the Markov chain jumps from state i to state j in one time step, i. e.,

$$q_{ij} := \Pr[X_{n+1} = j \mid X_n = i], \quad \forall n = 0, 1, \dots.$$

In particular,

$$\Pr(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = q_{x_0} \prod_{i=1}^{n} q_{x_{i-1},x_i},$$

where $q_{x_0} = \Pr(X_0 = x_0)$ is the initial distribution of the Markov chain. The n-step probability transition matrix $P^{(n)} = [q_{ij}^{(n)}]_{i,j\in\mathbb{I}}$ is determined by transition matrix P by the following relation:

https://doi.org/10.1515/9783110788105-015

$$P^{(n)} = \underbrace{P \circ P \circ \cdots \circ P}_{n\text{-fold matrix multiplication}},$$

where $q_{ij}^{(n)}$ is the probability that the Markov chain jumps from state i to j in n time steps, i. e., $q_{ij}^{(n)} = \Pr(X_n = j \mid X_0 = i), \forall i = 0, 1, 2, \ldots$.

The following is the celebrated formula for computing probability transition matrices.

Proposition 15.1.1 (Chapman–Kolmogorov). *For any $n \geq 0$, $m \geq 0$,*

$$P^{(n+m)} = P^{(n)} P^{(m)} = P^{(m)} P^{(n)},$$

where $P^{(n)} P^{(m)} = P^{(n)} \circ P^{(m)}$ and $P^{(m)} P^{(n)} = P^{(m)} \circ P^{(n)}$. That is, for all $i, j \in \mathbb{I}$,

$$q_{ij}^{(n+m)} = \sum_{k \in \mathbb{I}} q_{ik}^{(n)} q_{kj}^{(m)},$$

where $q_{ij}^{(n)}$ is the (i, j)-entry of the n-step probability transition matrix $P^{(n)} = [q_{ij}^{(n)}]_{i,j \in \mathbb{I}}$ for all $n \in \mathbb{N}$.

B. Stationary distribution

A time-homogeneous Markov chain $\{X_n\}_{n=0}^{+\infty}$ at time n is characterized by its distribution $\gamma^{(n)} = (\gamma_i^{(n)}, i \in \mathbb{I})$ (which is treated as a row vector), where $\gamma_i^{(n)} = \Pr(X_n = i)$ and $\gamma^{(n+1)} = \gamma^{(n)} P$, i. e., $\gamma_j^{(n+1)} = \sum_{i \in \mathbb{I}} \gamma_i^{(n)} q_{ij}$, for all $j \in \mathbb{I}$.

We have the following definition of *stationary distribution*:

– A distribution $\gamma^* = (\gamma_i^*, i \in \mathbb{I})$ is said to be a stationary distribution for the Markov chain $\{X_n\}_{n=0}^{+\infty}$ if $\gamma^* = \gamma^* P$, i. e., $\gamma_j^* = \sum_{i \in \mathbb{I}} \gamma_i^* q_{ij}$, for all $j \in \mathbb{I}$.

C. Classification of states

We list here a set of basic definitions (see Norris [117]).

– State j is said to be accessible from state i, denoted by the notation $i \to j$, if $q_{ij}^{(n)} > 0$ for some $n \geq 1$.
– State i and j communicate, denoted by $i \leftrightarrow j$, if both j is accessible from i and i is accessible from j. Note that the binary communication relation "$\leftrightarrow$" is an equivalence relation: reflexivity: i always communicates with i (by definition); symmetry: if i communicates with j, then j communicates with i (also by definition); transitivity: if i communicates with j and j communicates with k, then i communicates with k.
– Two states that communicate are said to belong to the same equivalence class, and the state space $\mathbb{I}$ is divided into a certain number of such classes.
– The Markov chain $\{X_n\}_{n=0}^{+\infty}$ is said to be *irreducible* if there is only one equivalence class (i. e., all states in $\mathbb{I}$ communicate with each other).

– A state i is absorbing if $q_{ii} = 1$.
– A state i is periodic with period L if L is the smallest integer $n \geq 1$ such that $q_{ii}^{(n)} > 0$. In case $L = 1$, the state is said to be aperiodic.
– It can be shown that if a state i is periodic with period L, then all states in the same class are periodic with the same period L, in which case the whole class is periodic with period L.

D. Ergodic Markov chains

A Markov chain $\{X_n\}_{n=0}^{+\infty}$ is said to be ergodic if there exists a positive integer T such that $q_{ij}^{(n)} > 0$ for all pairs $i, j \in \mathbb{I}$ and for all $n > T$.

For a Markov chain $\{X_n\}_{n=0}^{+\infty}$ to be ergodic, two technical conditions are required of its states and the nonzero transition probabilities; these conditions are known as irreducibility and aperiodicity. Informally, the first ensures that there is a sequence of transitions of nonzero probability from any state to any other, while the latter ensures that the states are not partitioned into sets such that all state transitions occur cyclically from one set to another.

A proof of the following theorem can be found in Norris [117].

Theorem 15.1.2. *For any ergodic Markov chain, there is a unique stationary probability vector* $\gamma^* = (\gamma_i^*, i \in \mathbb{I})$ *that is the principal left eigenvector of* P*, such that if* $\eta(i, n)$ *is the number of visits to state i in n steps, then*

$$\lim_{n \to +\infty} \frac{\eta(i, n)}{n} = \pi(i), \tag{15.1}$$

where $\pi(i) > 0$ *is the steady-state probability for state i.*

15.2 Constructions of Markov memory models

The presentation of the remainder of this chapter is largely based on results obtained by Datta and Dorlas [31–34], Dorlas and Morgan [42], and Rybar and Ziman [136].

In this section, we consider quantum channels with Markovian memory as first introduced by Macchiavello and Palma [111].

Let there be given a homogeneous Markov chain $\{X_n\}_{n=0}^{+\infty}$ defined on a probability space $(\Omega, \Sigma, \Pr)$ with a finite state space $\mathbb{I}$, (one-step) transition probabilities $P = [q_{ij}]_{i,j \in \mathbb{I}}$, i. e., $q_{ij} = \Pr(X_{n+1} = j \mid X_n = i)$ and the initial distribution $q_{i_0} = \Pr(X_0 = i_0)$. Let $\{\gamma_i\}_{i \in \mathbb{I}}$ be a stationary (or invariant) distribution for this chain, i. e.,

$$\gamma_j = \sum_{i \in \mathbb{I}} \gamma_i q_{ij}, \quad j \in \mathbb{I}. \tag{15.2}$$

To construct a quantum channel with Markovian memory, let $\Phi_i : \mathfrak{B}(\mathbb{H}_A) \to \mathfrak{B}(\mathbb{H}_B)$ be given completely positive trace-preserving (CPTP) maps for each $i \in \mathbb{I}$, where $\mathbb{H}_A$

and $\mathbb{H}_B$ are assumed to be finite-dimensional Hilbert spaces for illustration simplicity. However, many of the results presented in this chapter hold or can be easily extended to infinite-dimensional Hilbert spaces. The family of channels $\{\Phi_i\}_{i\in\mathbb{I}}$ will be referred to as component channels corresponding to the Markov chain.

In the following, we adapt the notation and terminologies introduced in Section 14.3 and consider the tensor product algebras $\mathfrak{A}^{(n)} = \mathfrak{B}(\mathbb{H}_A^{\otimes n})$ and the infinite tensor product C^*-algebra obtained as the strong closure

$$\mathfrak{A}^{(\infty)} = \overline{\bigcup_{n=1}^{\infty} \mathfrak{A}^{(n)}}^{\|\cdot\|_\infty}, \tag{15.3}$$

where we embed $\mathfrak{A}^{(n)}$ into $\mathfrak{A}^{(n+1)}$ in the obvious way and the closure is with respect to the operator norm $\|\cdot\|_\infty$. Similarly, we define $\mathfrak{B}^{(n)} = \mathfrak{B}(\mathbb{H}_B^{\otimes n})$ and $\mathfrak{B}^{(\infty)}$. A state on an algebra $\mathfrak{A}^{(\infty)}$ is a positive linear functional ω on $\mathfrak{A}^{(\infty)}$ with $\omega(\mathbf{I}) = 1$, where $\mathbf{I}$ denotes the identity operator on $\mathfrak{A}^{(\infty)}$. If $\mathfrak{A}^{(\infty)}$ is finite-dimensional, then there exists a density matrix $\rho_\omega \in \mathfrak{A}_\infty$ such that $\omega(\mathbf{A}) = \mathrm{tr}[\rho_\omega \mathbf{A}]$, for any $\mathbf{A} \in \mathfrak{A}^{(\infty)}$. We denote the quantum states on $\mathfrak{A}^{(\infty)}$ by $\mathcal{S}(\mathfrak{A}^{(\infty)})$ and those on $\mathfrak{A}^{(n)}$ by $\mathcal{S}(\mathfrak{A}^{(n)})$, etc. We now define a quantum channel with Markovian-correlated noise by the CPTP map $\Phi^{(\infty)} : \mathcal{S}(\mathfrak{A}^{(\infty)}) \to \mathcal{S}(\mathfrak{B}^{(\infty)})$ on the states of $\mathfrak{A}^{(\infty)}$ by

$$\Phi^{(\infty)}(\omega)(\mathbf{A}) = \sum_{i_1,i_2,\dots,i_n\in\mathbb{I}} \gamma_{i_1} q_{i_1 i_2} \cdots q_{i_{n-1} i_n} \,\mathrm{tr}\big[(\Phi_{i_1} \otimes \cdots \otimes \Phi_{i_n})(\rho_{\omega_n})\mathbf{A}\big] \tag{15.4}$$

for $\mathbf{A} \in \mathfrak{B}^{(n)}$. Here, ω_n is the restriction of ω to $\mathfrak{A}^{(n)}$ and ρ_{ω_n}, its density matrix. It is easily seen, using the property 15.2, that this definition is consistent and defines a CPTP map on the states of $\mathfrak{A}^{(\infty)}$, and moreover, that it is translation-invariant (stationary). We denote the transpose action of the restriction of $\Phi^{(\infty)}$ to $\mathcal{S}(\mathfrak{A}^{(n)})$ by $\Phi^{(n)} : \mathfrak{B}(\mathbb{H}_A^{\otimes n}) \to \mathfrak{B}(\mathbb{H}_B^{\otimes n})$, i. e.,

$$\mathrm{tr}\big[\Phi^{(n)}(\rho_\omega)\mathbf{A}\big] = \Phi^{(\infty)}(\omega)(\mathbf{A}),$$

for a density matrix $\rho_\omega \in \mathfrak{B}(\mathbb{H}_A^{\otimes n})$, $\omega \in \mathcal{S}(\mathfrak{A}^{(n)})$. Thus, $\Phi^{(n)}$, the n-use of the channel with Markovian memory can be expressed as

$$\Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A}) = \sum_{i_1,i_2,\dots,i_n\in\mathbb{I}} q_{i_1} q_{i_1 i_2} \cdots q_{i_{n-1} i_n} (\Phi_{i_1} \otimes \cdots \otimes \Phi_{i_n})(\rho^{(n)}_{\Lambda_n,A}), \tag{15.5}$$

for the codeword $\rho^{(n)}_{\Lambda_n,A} \in \mathcal{S}(\mathfrak{A}^{(n)})$ of the classical data $\Lambda_n = \{\lambda_1, \lambda_2, \dots, \lambda_{M_n}\}$ of length n and size M_n.

We can also write $\Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A})$ defined in (15.5) in terms of the Markov chain $\{X_n\}_{n=1}^{+\infty}$ and the family of component channels $\{\Phi_i\}_{i\in\mathbb{I}}$ as follows:

$$\Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A}) = \mathbb{E}\big[(\Phi_{X_1} \otimes \cdots \otimes \Phi_{X_n})(\rho^{(n)}_{\Lambda_n,A})\big], \tag{15.6}$$

where $\mathbb{E}[\cdots]$ is the expectation of $[\cdots]$ associated with the probability space $(\Omega, \Sigma, \Pr)$ of the Markov chain $\{X_n\}_{n=0}^{+\infty}$.

The CPTP map $\Phi^{(\infty)} : \mathcal{S}(\mathfrak{A}^{(\infty)}) \to \mathcal{S}(\mathfrak{B}^{(\infty)})$ defined in (15.5) will be referred to as a quantum channel with Markovian correlated noise or simply quantum channel with Markov memory.

Writing (15.5) in the form of Kraus representation, we have

$$\Phi^{(n)}(\rho_{\Lambda_n,A}^{(n)}) = \sum_{i_1,i_2,\ldots,i_n\in\mathbb{I}} q_{i_1}q_{i_1,i_2}\cdots q_{i_{n-1},i_n}(\mathbf{A}_{i_n}\cdots\mathbf{A}_{i_1})\rho_{\Lambda_n,A}^{(n)}(\mathbf{A}_{i_0}^*\cdots\mathbf{A}_{i_n}^*). \tag{15.7}$$

15.3 Channels with ergodic Markovian memory

In this section, we assume that the underlying Markov chain $\{X_n\}_{n=0}^{+\infty}$ is aperiodic and irreducible so that, in particular, the invariant distribution, $\{\gamma_i\}_{i\in\mathbb{I}}$, is unique. It is well known in Norris [117] that the corresponding Markov chain is ergodic, and consequently, the output states of the channel are also ergodic. In this case, the Markov chain satisfies the property of convergence to equilibrium, i. e.,

$$\lim_{n\to+\infty} q_{ij}^{(n)} = \gamma_j, \quad \forall i \in \mathbb{I},$$

where $q_{ij}^{(n)}$ denotes the n-step transition probability from the state i to the state j, $(i,j \in \mathbb{I})$. This implies that the correlation in the noise, acting on successive inputs to the channel, dies out after a sufficiently large number of uses of the channel. Hence, in this case the family of n-uses of the channel $\{\Phi^{(n)}\}_{n=1}^{+\infty}$ defined in (15.6) belongs to the class of channels introduced and studied by Kretschmann and Werner [100], and referred to as forgetful channels treated in Section 14.4.

15.3.1 Classical capacity

To explore the classical capacity of the channel with ergodic Markov memory, we recall (see Definition 12.1.2), the channel output Holevo χ-quantity as follows.

For each $n \in \mathbb{N}$, suppose that $\{p_j^{(n)}, \rho_j^{(n)}\}_{j=1}^{J(n)}$ is an ensemble of states given by density operators $\rho_j^{(n)}$ on $\mathbb{H}_A^{\otimes n}$ with probabilities $p_j^{(n)} > 0$, $\sum_{j=1}^{J(n)} p_j^{(n)} = 1$, where $J(n)$ is a positive integer that depends on n. In this case, the Holevo quantity for the channel restricted to $\mathfrak{A}^{(n)} = \mathfrak{B}(\mathbb{H}_A^{\otimes n})$ is given by

$$\begin{aligned}\chi(\{p_j^{(n)}, \Phi^{(n)}(\rho_j^{(n)})\}_{j=1}^{J(n)}) \\ = H\left(\sum_{j=1}^{J(n)} p_j^{(n)}\Phi^{(n)}(\rho_j^{(n)})\right) - \sum_{j=1}^{J(n)} p_j^{(n)} H(\Phi^{(n)}(\rho_j^{(n)})).\end{aligned} \tag{15.8}$$

Lemma 15.3.1. *Let $\Phi^{(n)}$ be the n-use of the quantum channel with ergodic Markov memory described in* (15.5). *Then the following limit exists:*

$$\chi^*(\Phi) = \lim_{n\to+\infty} \frac{1}{n} \sup_{\{p_j^{(n)},\rho_j^{(n)}\}} \chi(\{p_j^{(n)}, \Phi^{(n)}(\rho_j^{(n)})\}) \tag{15.9}$$

We first note that the limit (15.10) in the following result exists by Corollary 15.4.9, where the classical capacity of channels with ergodic Markov memory is stated.

Theorem 15.3.2. *The classical capacity of a quantum channel with memory, $\Phi^{(\infty)}$, defined by* (15.4), *where the underlying Markov chain is aperiodic and irreducible, is given by*

$$\chi^*(\Phi^{(\infty)}) = \lim_{n\to+\infty} \frac{1}{n} \sup_{\{p_j^{(n)},\rho_j^{(n)}\}} \chi(\{p_j^{(n)}, \Phi^{(n)}(\rho_j^{(n)})\}). \tag{15.10}$$

First, we note that the limit defining $\chi^*(\Phi^{(\infty)})$ in (15.10) exists by Lemma 15.3.1. The proof of Theorem 15.3.2 consists of the following two parts:

(i) Proposition 15.3.3, the direct part of Theorem 15.3.2, proves that for any rate $R < \chi^*(\Phi)$ is achievable.

(ii) The weak converse part of Theorem 15.3.2 is a special case of Proposition, which proves that it is impossible for the sender to transmit classical messages reliably to the receiver through the channel $\Phi^{(\infty)}$ at a rate $R > \chi^*(\Phi^{(\infty)})$.

The following quantum version of the Feinstein lemma is due originally to Datta and Dorlas [33].

Proposition 15.3.3 (Quantum version of the Feinstein lemma). *Let $\Phi^{(\infty)}$ be a quantum channel with Markov memory defined by* (15.4), *where the underlined Markov chain is ergodic. Let $\chi^* = \chi^*(\Phi^{(\infty)})$ be given in* (15.10). *Then, given $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ there exist at least $N \geq 2^{n(\chi^*-\epsilon)}$ states with states $\tilde{\rho}_1^{(n)}, \ldots \tilde{\rho}_N^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n})$, and positive operators $\mathbf{D}_1^{(n)}, \ldots, \mathbf{D}_N^{(n)} \in \mathcal{S}(\mathbb{H}_B^{\otimes n})$ such that $\sum_{k=1}^N \mathbf{D}_k^{(n)} \leq \mathbf{I}^{(n)}$ (where $\mathbf{I}^{(n)}$ is the identity operator on $\mathbb{H}_B^{\otimes n}$) and*

$$\mathrm{tr}[\Phi^{(n)}(\tilde{\rho}_k^{(n)})\mathbf{D}_k^{(n)}] > 1 - \epsilon. \tag{15.11}$$

Proof. By the definition of χ^* given in (15.10), for the given $\epsilon > 0$, we can choose $l_0 \in \mathbb{N}$ so large that

$$\left| \frac{1}{l_0} \sup_{\{p_j^{(l_0)},\rho_j^{(l_0)}\}} \chi(p_j^{(l_0)}, \Phi^{(l_0)}(\rho_j^{(l_0)})) - \chi^* \right| < \frac{\epsilon}{6}. \tag{15.12}$$

Assume that the supremum in (15.12) is attained at an ensemble $\{p_j^{(l_0)}, \rho_j^{(l_0)}\}_{j=1}^J$ for a finite J. Denote for $m \in \mathbb{N}$,

$$\bar{\sigma}_{ml_0} = \Phi^{(ml_0)}((\bar{\rho}^{(l_0)})^{\otimes m}), \tag{15.13}$$

where $\bar{\rho}^{(l_0)} = \sum_{j=1}^{J} p_j^{(l_0)} \rho_j^{(l_0)}$. These states $\{\bar{\sigma}_{ml_0}\}_{m=1}^{+\infty}$ form a compatible system of states on $\{\mathfrak{B}^{(ml_0)}\}_{m=1}^{+\infty}$ (where $\mathfrak{B}^{(ml_0)} = \mathfrak{B}(\mathbb{H}_B^{\otimes ml_0})$), and hence exists a state $\bar{\phi}^{(\infty)}$ on $\mathfrak{B}^{(\infty)}$ such that

$$\bar{\phi}^{(\infty)}(\mathbf{A}) = \mathrm{tr}[\bar{\sigma}_{ml_0}\mathbf{A}], \quad \forall \mathbf{A} \in \mathfrak{B}^{(ml_0)}. \tag{15.14}$$

This state is clearly l_0-periodic, i. e., invariant under translations over multiples of l_0. Therefore, the following mean entropy $H_M(\bar{\phi}^{(\infty)})$ exists:

$$H_M(\bar{\phi}^{(\infty)}) := \lim_{m\to+\infty} \frac{1}{m} H(\bar{\sigma}_{ml_0}) = \inf_{m\in\mathbb{N}} \frac{1}{m} H(\bar{\sigma}_{ml_0}). \tag{15.15}$$

The following lemma shows that for l_0 sufficiently large, the mean entropy, $H_M(\bar{\phi}^{(\infty)})$, is close to the von Neumann entropy of the average output of l_0 uses of the channel, $H(\bar{\sigma}_{l_0}) = H(\Phi^{(l_0)}(\bar{\rho}^{(l_0)}))$.

Precisely, we have the following.

Lemma 15.3.4. *Given $\epsilon > 0$, there exists $L > 0$ such that for $l_0 \geq L$,*

$$\left| \frac{1}{l_0} H_M(\bar{\phi}^{(\infty)}) - \frac{1}{l_0} H(\Phi^{(l_0)}(\bar{\rho}^{(l_0)})) \right| < \frac{\epsilon}{8}, \tag{15.16}$$

where $\bar{\phi}^{(\infty)}$ is given by (15.14) *and $\bar{\rho}^{(l_0)} = \frac{1}{J}\sum_{j=1}^{J} \rho_j^{(l_0)}$.*

The proof is similar to (15.12) and is therefore omitted here.

Henceforth, the natural number l_0 is assumed to be fixed to a value such that Lemma 15.3.4 and (15.12) hold. For notational simplicity, explicit dependence on l_0 is often suppressed.

The proof of Proposition 15.3.3 requires the sequence of lemmas given below.

Lemma 15.3.5. *The state $\bar{\phi}^{(\infty)}$ is strongly clustering, and hence completely ergodic for l_0-shifts, i. e., for any $\mathbf{A}, \mathbf{B} \in \mathfrak{B}^{(ml_0)}$,*

$$\lim_{k\to+\infty} \bar{\phi}_\infty(\mathbf{A}\tau^{kl_0}(\mathbf{B})) = \mathrm{tr}[\bar{\sigma}_{ml_0}\mathbf{A}]\,\mathrm{tr}[\bar{\sigma}_{ml_0}\mathbf{B}], \tag{15.17}$$

where τ^{ml_0} is the shift operator defined in Subsection 14.3.1.

Proof. The proof is standard and relies on the fact that the expectations of $\mathbf{A}$ and $\mathbf{B}$ in the state $\bar{\phi}^{(\infty)}$ decouple as their supports are separated by a sufficiently large distance. This is because

$$\lim_{k\to+\infty} q_{i_1 i_2} \cdots q_{i_{k-1} i_k} g(i_k) = \gamma_i g(i)$$

for any function $g(i)$, since the Markov chain is irreducible and aperiodic. This proves the lemma. □

We also use the following lemma.

Lemma 15.3.6. *For any $\delta > 0$, there exists $m_1 \in \mathbb{N}$ such that for all $m \geq m_1$ there exists a subspace $\mathbb{T}_\epsilon^{(m)} \subset \mathbb{H}_B^{\otimes l_0 m}$ with projection $\bar{P}_{ml_0}$ such that*

$$\bar{\mathbf{P}}_{ml_0} \bar{\sigma}_{ml_0} \bar{\mathbf{P}}_{ml_0} \leq 2^{-m[H_M(\bar{\phi}^{(\infty)}) - \frac{\epsilon}{4}]} \mathbf{1}^{(ml_0)} \tag{15.18}$$

and

$$\operatorname{tr}[\bar{\sigma}^{(ml_0)} \bar{\mathbf{P}}^{(ml_0)}] > 1 - \delta^2, \tag{15.19}$$

where $\mathbf{1}^{(ml_0)}$ is the identity operator on the space $\mathbb{H}_B^{\otimes ml_0}$.

Proof. Let l_1 be so large that

$$H_M(\bar{\phi}^{(\infty)}) \leq \frac{1}{l_1} H(\bar{\sigma}^{(l_1)}) < H_M(\bar{\phi}^{(\infty)}) + \frac{\epsilon}{8}.$$

Let $\Omega = \{\lambda_k\}$ denote the spectrum of $\bar{\sigma}^{(l_1)} = \Phi^{(l_1)}(\bar{\rho}^{(l_1)})$, and let π_k be the projection onto the eigenvector with eigenvalue λ_k. For any $r > 0$ and $C \subset \Omega^{\times r}$, put

$$\pi_C = \sum_{(\lambda_{k_1}, \lambda_{k_2}, \ldots, \lambda_{k_r}) \subset C} \pi_{k_1} \otimes \pi_{k_2} \otimes \cdots \otimes \pi_{k_r},$$

and define the probability measures ν_r on $\Omega^{\times r}$ and ν_∞ on $\Omega^{\mathbb{N}}$ by

$$\nu_r(C) = \operatorname{tr}[\Phi^{(rl_0 l_1)}(\bar{\rho}_{l_0}^{\otimes (rl_1)}) \pi_C] \quad \text{and} \quad \nu_\infty(C) = \bar{\phi}^{(\infty)}(\pi_C).$$

We state McMillan's theorem [112] below: For each source symbol from the alphabet,

$$S = \{s_1, s_2, \ldots, s_r\}$$

that is encoded into a uniquely decodable code over an alphabet of size n with codeword lengths,

$$\theta_1, \theta_2, \ldots, \theta_n,$$

then

$$\sum_{i=1}^{n} r^{-\theta_i} \leq 1.$$

Conversely, for a given set of natural numbers,

$$\ell_1, \ell_2, \ldots, \ell_n$$

satisfying the above inequality, there exists a uniquely decodable code over an alphabet of size r with those codeword lengths. By Lemma 15.3.5, ν_∞ is ergodic, and there exists a typical set

$$T_\epsilon^{(r)} = \{(\lambda_{k_1}, \ldots, \lambda_{k_r}) \in \Omega^r \mid 2^{-r(h_{\mathrm{KS}}(\nu_\infty)+\frac{\epsilon}{8})} \\ \leq \nu_r(\{(\lambda_{k_1}, \ldots, \lambda_{k_r})\}) \leq 2^{-r(h_{\mathrm{KS}}(\nu_\infty)-\frac{\epsilon}{8})}\} \tag{15.20}$$

satisfying $\nu_r(\mathbb{T}_\epsilon^{(r)}) > 1 - \delta^2$ for r large enough, where $h_{\mathrm{KS}}(\nu_\infty)$ denotes the Kolmogorov–Sinai entropy (see Billingsley [10] for a definition). Now,

$$h_{\mathrm{KS}}(\nu_\infty) = \inf_r \frac{1}{r} S(\nu_r) \leq S(\nu_1) = H(\bar{\sigma}^{(l_1)}) < l_1\left(H_M(\bar{\phi}_\infty) + \frac{\epsilon}{8}\right), \tag{15.21}$$

where $S(\nu)$ denotes the Shannon entropy corresponding to the measure ν. On the other hand,

$$h_{\mathrm{KS}}(\nu_\infty) \geq H_M(\bar{\phi}_\infty) \tag{15.22}$$

because, by positivity of the relative entropy,

$$\begin{aligned} H(\bar{\sigma}^{(rl_1)}) &= -\operatorname{tr}[\bar{\sigma}^{(rl_1)} \log \sigma^{(rl_1)}] \\ &\leq -\operatorname{tr}\left[\bar{\sigma}^{(rl_1)} \log\left(\bigoplus_{k_1,\ldots,k_r} \operatorname{tr}[\bar{\sigma}^{(rl_1)}(\pi_{k_1} \otimes \cdots \otimes \pi_{k_r})](\pi_{k_1} \otimes \cdots \otimes \pi_{k_r})\right)\right] \\ &= -\sum_{k_1,\ldots,k_r} \operatorname{tr}[\bar{\sigma}^{(rl_1)}(\pi_{k_1} \otimes \cdots \otimes \pi_{k_r})] \log(\operatorname{tr}[\bar{\sigma}^{(rl_1)}(\pi_{k_1} \otimes \cdots \otimes \pi_{k_r})]) = S(\nu_r). \end{aligned}$$

For arbitrary m, let $r = [m/l_1]$ and define

$$\pi_{\mathbf{k}}^{(m)} = \pi_{k_1} \otimes \cdots \otimes \pi_{k_r} \otimes \mathbf{1} \in \mathfrak{B}(\mathbb{H}_B^{\otimes ml_0}), \quad \mathbf{k} = (k_1, k_2, \ldots, k_r).$$

Let

$$\bar{T}_\epsilon^{(m)} = \{\mathbf{k} \mid (\lambda_{k_1}, \ldots, \lambda_{k_r}) \in T_\epsilon^{(r)}\},$$

and define

$$\mathbb{T}_\epsilon^{(m)} = \bigoplus_{\mathbf{k} \in \bar{T}_\epsilon^{(m)}} \pi_{\mathbf{k}}^{(m)}(\mathbb{H}_B^{\otimes ml_0}).$$

Clearly,

$$\bar{\phi}_\infty\Big(\bigoplus_{\mathbf{k}\in\bar{T}_{\mathbf{k}}^{(m)}} \pi_{\mathbf{k}}^{(m)}\Big) = \operatorname{tr}[\bar{\sigma}^{(rl_1)}\pi_{\bar{T}_{\mathbf{k}}^{(r)}}] = \nu_r(T_\epsilon^{(r)}) > 1-\delta^2.$$

Moreover, if $\vec{k}\in\bar{T}_\epsilon^{(m)}$, it follows from (15.21) and (15.22) that

$$\frac{1}{m}\log\nu(\{(\lambda_{k_1},\ldots,\lambda_{k_r})\}) \le -\frac{rl_1}{m}\Big(H_M(\bar{\phi}_\infty) - \frac{1}{l_1}\frac{\epsilon}{8}\Big),$$

and

$$\frac{1}{m}\log\nu(\{(\lambda_{k_1},\ldots,\lambda_{k_r})\}) \ge -\frac{rl_1}{m}\Big(H_M(\bar{\phi}_\infty) + \Big(1+\frac{1}{l_1}\Big)\frac{\epsilon}{8}\Big).$$

Taking $l_1 > 3$ and m large enough, we obtain

$$\Big|\frac{1}{m}\log\nu(\{(\lambda_{k_1},\ldots,\lambda_{k_r})\}) + \frac{rl_1}{m}(H_M(\bar{\phi}_\infty))\Big| < \frac{\epsilon}{6}.$$

Now let

$$\bar{\mathbf{P}}_{ml_0} = \bigoplus_{\vec{k}\in\bar{T}_\epsilon^{(m)}} \pi_{\vec{k}}^{(m)}$$

and assume that l_1 is so large that $\epsilon l_1/12 > -\log\gamma_{\min}$, where $\gamma_{\min} = \wedge_{i\in I}\gamma_i$.

Note that $\gamma_{\min} > 0$. Define

$$\bar{\sigma}(l,l') = \sum_{j_1,\ldots,j_l=1}^{J} p_{\vec{j}}^{(l)} \sum_{i_2,\ldots,i_{l-1}} \gamma_i q_{ii_2}q_{i_2i_3}\cdots q_{i_{l-1}i'}\Phi_i(\rho_{j_1})\otimes\cdots\otimes\Phi_{i'}(\rho_{j_l}),$$

where $\vec{j} = (j_1, j_2, \ldots, j_l)$. Then we can write as in the proof of Lemma 15.3.4,

$$\begin{aligned}\bar{\sigma}_{ml_0} = &\sum_{i_1,i_2,\ldots,i_{2r+1}} \frac{q_{i_2i_3}}{\gamma_{i_3}}\frac{q_{i_4i_5}}{\gamma_{i_5}}\cdots\frac{q_{i_{2r}i_{2r+1}}}{\gamma_{i_{2r+1}}}\\ &\times\bar{\sigma}_{l_1}(i_1,i_2)\otimes\cdots\otimes\bar{\sigma}_{l_1}(i_{2r-1},2r)\otimes\bar{\sigma}^{(m-rl_1)}(i_{2r-1},2r).\end{aligned}$$

Using the positivity of the transition probabilities, we have

$$\bar{\mathbf{P}}_{ml_0}\bar{\sigma}_{ml_0}\bar{\mathbf{P}}_{ml_0} \le 2^{-m[H_M(\bar{\phi}_\infty)-\frac{\epsilon}{4}]}\mathbf{1}^{(ml_0)}.$$

By the fact that π_k is an eigenprojection of $\bar{\sigma}^{(l_1)}$, we then have

$$\bar{\mathbf{P}}_{ml_0}\sigma_{ml_0}\bar{\mathbf{P}}_{ml_0} \le \gamma^{-r}2^{-m[H_M(\bar{\phi}_\infty)-\frac{\epsilon}{6}]}\mathbf{1}^{(ml_0)}.$$

But $\gamma^{-r} < 2^{-m\epsilon/12}$ by the above assumption. This proves the lemma. □

Let $J \in \mathbb{N}$, and let

$$\mathbb{J}^n = \{(j_1, j_2, \ldots, j_n) \mid j_i = 1, \ldots, J, \text{ for } i = 1, 2, \ldots, n\}.$$

Let $\sigma_j \in \mathcal{S}(\mathbb{H}_B)$ for each $j = 1, \ldots, J$ with eigenvalue λ_{jk}, where $k = 1, 2, \ldots, d$, in which $d = \dim(\mathbb{H}_B)$ is the dimension of the Hillbert space $\mathbb{H}_B$.

For any given $\vec{j}$, define

$$\mathbf{V}_{\vec{j}}^{(n)} = \left(\bar{\Pi}_n - \sum_{k=1}^{N} \mathbf{D}_k^{(n)}\right)^{1/2} \bar{\Pi}_n \Pi_{\vec{j}}^{(n)} \bar{\Pi}_n \left(\bar{\Pi}_n - \sum_{k=1}^{N} \mathbf{D}_k^{(n)}\right)^{1/2}. \tag{15.23}$$

Clearly, $\mathbf{V}_{\mathbf{j}}^{(n)} \le \mathbf{P}_n - \sum_{k=1}^{N} \mathbf{D}_k^{(n)}$. We also have the following result.

Lemma 15.3.7. *Given $\delta > 0$, there exists $m_2 \in \mathbb{N}$ such that for all $m \ge m_2$ there exist, for all $\vec{j} = (j_1, \ldots, j_m) \in \{1, \ldots, J\}^{\times m}$, one-dimensional subspaces $\mathbb{T}_{\vec{j},\vec{k}}^{(m)}$ of $\mathbb{H}_B^{\otimes m l_0}$ (indexed by $\vec{k}$ in some set $T_{\vec{j},\epsilon}^{(m)}$) with projections $\pi_{\vec{j},\vec{K}}^{(ml_0)}$ in the $\vec{j}$th component of Λ_{ml_0}, such that for all $\vec{k} \in T_{\vec{j},\epsilon}^{(m)}$,*

$$\left|\frac{1}{m} \log(\omega_{\vec{j},\vec{k}}^{(ml_0)}) + H_M(\{p_j^{(l_0)}, \rho_j^{(l_0)}\})\right| < \frac{\epsilon}{4},$$

where $\omega_{\vec{j},\vec{k}}^{(ml_0)} = \mathrm{tr}[\Phi^{(ml_0)}(\rho_{\vec{j},\vec{k}}^{(ml_0)})]$ and

$$\psi_\infty\left(\bigoplus_{\vec{j}} \bigoplus_{\vec{k} \in T_{\vec{j},\epsilon}^{(m)}} \pi_{\vec{j},\vec{k}}^{(ml_0)}\right) > 1 - \delta^2.$$

Proof. In the following, we suppress the dependence on l_0. We follow the argument in the proof of Lemma 15.3.6. Fix $l \ge 12$ large enough so that

$$\frac{1}{l} H(\Sigma_{ll_0}) < H_M(\psi_\infty) - \frac{\epsilon}{12}. \tag{15.24}$$

Let $\mathcal{Y}^{(l)}$ be the spectrum of $\sigma^{(ll_0)} = \Phi^{(ll_0)}(\rho_{j_1}^{(l_0)} \otimes \rho_{j_2}^{(l_0)} \otimes \cdots \otimes \rho_{j_l}^{(l_0)})$. Note that Σ_{ll_0} can be represented as a block-diagonal matrix in $\bigoplus_{j_1,\ldots,j_l=1}^{J} \mathbb{H}_B^{\otimes ll_0}$ with spectrum consisting of eigenvalues $\nu_{\vec{j},k} = p_{\vec{j}}^{(l)} \alpha_{\vec{j},k}$ with $\vec{j} \in \{1, 2, \ldots, J\}^l$, $k = 1, 2, \ldots, (\dim(\mathbb{H}_B^{\otimes l_0}))^l$, and $\alpha_{\vec{j},k}$ being the eigenvalues of $\sigma_{\vec{j}}^{(ll_0)}$. Let

$$\mathcal{Y}_l = \bigcup_{\vec{j} \in \{1,2,\ldots,J\}^l} \mathcal{Y}_{\vec{j}}^{(l)}.$$

For each $s \in \mathbb{N}$, we now define measures μ_s on $(\mathcal{Y}_l)^s$ by

$$\mu_s(C) = \sum_{\vec{j}\in\{1,2,\ldots,J\}^{sl}} p_{\vec{j}}^{(sl)} \operatorname{tr}[\sigma_{\vec{j}}^{(sl)} \mathbf{q}_C^{(s)}], \quad C \subset (\mathcal{Y}_l)^s,$$

and

$$\mathbf{q}_C^{(s)} = \sum_{(\lambda_{\vec{j}_1,k_1},\ldots,\lambda_{\vec{j}_s,k_s})\in C} \pi_{\vec{j}_1,k_1} \otimes \cdots \otimes \pi_{\vec{j}_s,k_s}, \quad \vec{j} = (\vec{j}_1,\ldots,\vec{j}_s),$$

where $\pi_{\vec{j},k}$ denotes the projection onto the kth eigenvector of $\sigma_{\vec{j}}^{(l)}$. We also define the projective limit μ_∞ on $\mathcal{Y}_l^{\mathbb{N}}$ by

$$\mu_\infty(C) = \mu_s(C) = \psi_\infty(\mathbf{q}_C^{(s)}),$$

for a cylinder set $C \in (\mathcal{Y}_l)^s$. It follows that μ_∞ is ergodic. Define typical sets

$$\tilde{T}_{\vec{j},\epsilon}^{(s)} = \{(\lambda_{\vec{j}_1,k_1},\ldots,\lambda_{\vec{j}_s,k_s}) \in \mathcal{Y}_l^s \mid 2^{-s(h_{\mathrm{KS}}(\mu_\infty)+\frac{\epsilon}{12})} \le \mu_s(\{(\lambda_{\vec{j}_1,k_1},\ldots,\lambda_{\vec{j}_s,k_s})\}), \le 2^{-s(h_{\mathrm{KS}}(\mu_\infty)-\frac{\epsilon}{12})}\}$$

where $h_{\mathrm{KS}}(\nu_\infty)$ denotes the Kolmogorov–Sinai entropy of μ_∞ By the McMillan theorem,

$$\mu_s\Big(\big(\bigcup\big)_{\vec{j}} \tilde{T}_{\vec{j},s}^{(s)}\Big) > 1 - \frac{1}{2}\delta^2$$

for s large enough. Now,

$$h_{\mathrm{KS}}(\mu_\infty) = \inf_s \frac{1}{s} S(\mu_s) \le S(\mu_1) = H(\Sigma_l) = H(\bar{\sigma}^{(l_1)}) < l\Big(H_M(\bar{\psi}_\infty) + \frac{\epsilon}{12}\Big),$$

by (15.24), and on the other hand,

$$h_{\mathrm{KS}}(\mu_\infty) \ge l H_M(\psi_\infty)$$

by positivity of the relative entropy.

For arbitrary m, we argue as in Lemma 15.3.4, and let $s = [m/l]$. Writing, $m = sl + r$ and $\vec{j} = (j_1,\ldots,j_m) = (j_1,\ldots,j_s,j_0)$, we have

$$\pi^{(ml_0)})_{\vec{j},\vec{k}} = \pi_{\vec{j}_1,k_1} \otimes \cdots \otimes \pi_{\vec{j}_s,k_s} \otimes \pi_{\vec{j}_0}^{(r)},$$

where $\pi_{\vec{j}_0}^{(r)}$ is the projection of $\bigoplus_{j_1,\ldots,j_r=1}^{J} \mathbb{H}_B^{\otimes r}$ onto the j_0th summand. Let $\tilde{T}_{\vec{j},\epsilon}^{[m]} = \tilde{T}_{\vec{j}.\epsilon}^{(s)}$. Then

$$\begin{aligned}\psi_\infty\Big(\bigoplus_{\vec{j}\in\{1,2,\ldots,J\}^m} \bigoplus_{\vec{k}\in\tilde{T}_{\vec{k},\epsilon}^{(m)}} \pi_{\vec{j},\vec{k}}^{(ml_0)}\Big) &= \operatorname{tr}\Big[\sum_{sl}\Big(\bigoplus_{\vec{j}\in\{1,2,\ldots,J\}^{sl}} \mathbf{q}_{\tilde{T}_{\vec{j},\epsilon}^{(s)}}\Big)\Big] \\ &= \mu_s\Big(\bigcup_{\vec{j}} \tilde{T}_{\vec{j},\epsilon}^{(s)}\Big) > 1 - \frac{1}{2}\delta^2.\end{aligned}$$

Moreover, if $(\lambda_{\vec{j}_1,k_1},\dots,\lambda_{\vec{j}_s,k_s}) \in \tilde{T}^{[m]}_{\vec{j},\epsilon}$,

$$\frac{1}{m}\log\mu(\{(\lambda_{\vec{j}_1,k_1},\dots,\lambda_{\vec{j}_s k_s})\}) \le -\frac{sl}{m}\left(H_M(\psi_\infty) - \frac{1}{l}\frac{\epsilon}{12}\right),$$

and

$$\frac{1}{m}\log\mu(\{(\lambda_{\vec{j}_1,k_1},\dots,\lambda_{\vec{j}_s k_s})\}) \ge -\frac{sl}{m}\left(H_M(\psi_\infty) + \left(1+\frac{1}{l}\right)\frac{\epsilon}{12}\right).$$

Finally, define the typical set of indices $\vec{j}$:

$$T^{[m]}_\epsilon = \left\{\vec{j} \in \{1,2,\dots,J\}^m \mid 2^{-m(S(\{p_j\})+\frac{\epsilon}{12})} \le p^{(m)}_{\vec{j}} \le 2^{-m(S(\{p_j\})-\frac{\epsilon}{12})}\right\}$$

Then for m large enough, we obtain

$$\mathbb{P}^{\otimes m}[T^{(m)}_\epsilon] = 1 - \frac{1}{2}\delta^2,$$

if $\mathbb{P}$ denotes the probability with respect to the ensemble probabilities $\{p_j\}_{j=1}^J$. Defining

$$T^{(m)}_{\vec{j},\epsilon} = \begin{cases} T^{[m]}_{\vec{j},\epsilon} & \text{if } \vec{j} \in T^{[m]}_\epsilon \\ \emptyset & \text{if } \vec{j} \notin T^{[m]}_\epsilon, \end{cases}$$

we have for $(\lambda_{\vec{j}_1,k_1},\dots,\lambda_{\vec{j}_s k_s}) \in T^{(m)}_{\vec{j},\epsilon}$,

$$\begin{aligned}\frac{1}{m}\log(\lambda^{(m)}_{\vec{j},\vec{k}}) &= -\frac{1}{m}\log(\{(p^{(l)}_{\vec{j}_1},\dots,p^{(l)}_{\vec{j}_s})\}) + \frac{1}{m}\log\mu_s(\{(\lambda_{\vec{j}_1,k_1},\dots,\lambda_{\vec{j}_s k_s})\}) \\ &\le -\frac{sl}{m}\left(H_M(\psi_\infty) - \frac{1}{l}\frac{\epsilon}{12}\right) + S(\{p_j\}) + \frac{\epsilon}{12} \\ &\le -\bar{H}_M + \frac{\epsilon}{4}\end{aligned}$$

and

$$\begin{aligned}\frac{1}{m}\log(\lambda^{(m)}_{\vec{j}\vec{k}}) &= -\frac{1}{m}\log(\{(p^{(l)}_{\vec{j}_1},\dots,p^{(l)}_{\vec{j}_s})\}) + \frac{1}{m}\log\mu_s(\{(\lambda_{\vec{j}_1,k_1},\dots,\lambda_{\vec{j}_s k_s})\}) \\ &\ge -\left(\bar{H}_M + \frac{\epsilon}{4}\right)\end{aligned}$$

for m large enough. Moreover,

$$\psi_\infty\Big(\bigoplus_{\vec{j}\in T_\epsilon^{[m]}}\bigoplus_{\vec{k}\in T_{\vec{j},\epsilon}^{(m)}}\pi_{\vec{j},\vec{k}}^{(ml_0)}\Big) = \mathrm{tr}\Big[\Sigma_m\Big(\bigoplus_{\vec{j}\in T_\epsilon^{[m]}}\mathbf{q}_{\tilde{T}_{\vec{j},\epsilon}^{(s)}}\Big)\Big]$$

$$= \sum_{\vec{j}\in T_{\vec{j}}^{[m]}p_{\vec{j}}^{(m)}}\mathrm{tr}[\Phi^{(m)}(\rho_{\vec{j}}^{(m)})\mathbf{q}_{\tilde{T}_{\vec{j},\epsilon}^{(s)}}] = \mu_s\Big(\bigcup_{\vec{j}}\tilde{T}_{\vec{j},\epsilon}^{(s)}\Big) - \mathbb{P}^{\otimes m}[(T_\epsilon^{[m]})^c]$$

$$> 1-\delta^2.$$

This proves the lemma. □

We now continue the proof of Proposition 15.3.3. Consider the averaged von Neumann entropy $\bar{H} = \sum_{j=1}^J p_i H(\sigma_j)$, where $p_j \geq 0$ with $\sum_{j=1}^J p_j = 1$ and $H(\sigma_j) = -\sigma_j \log \sigma_j$ is the von Neumann entropy of the output state $\sigma_j \in \mathcal{S}(\mathbb{H}_B)$.

For each $n \in \mathbb{N}$, let $N = N(n)$ be the maximal number for which there exists states $\tilde{\rho}_1^{(n)}, \tilde{\rho}_2^{(n)}, \ldots, \tilde{\rho}_N^{(n)}$ on $\mathbb{H}_A^{\otimes n}$ of the tensor product form

$$\tilde{\rho}_k^{(n)} = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n,$$

and there exists positive operators $\mathbf{D}_1^{(n)}, \mathbf{D}_2^{(n)}, \ldots, \mathbf{D}_N^{(n)}$ on $\mathbb{H}_B^{\otimes n}$ such that, defining $\tilde{\sigma}_k^{(n)} = \Phi^{\otimes n}(\tilde{\rho}_k^{(n)})$, we have:

1. $\sum_{k=1}^N \mathbf{D}_k^{(n)} \leq \mathbf{P}_n$ and
2. $\mathrm{tr}[\tilde{\sigma}_k^{(n)}\mathbf{D}_k^{(n)}] > 1-\epsilon$ for each k, and
3. $\mathrm{tr}[\bar{\sigma}_n\mathbf{D}_k^{(n)}] \leq 2^{-n(H(\bar{\sigma})-\bar{H}-\frac{2}{3}\epsilon)}$ for each $k = 1, 2, \ldots, n$.

For any given $\vec{j}$, define

$$\mathbf{V}_{\vec{j}}^{(n)} = \Big(\bar{\Pi}_n - \sum_{k=1}^N \mathbf{D}_k^{(n)}\Big)^{1/2}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n\Big(\bar{\Pi}_n - \sum_{k=1}^N \mathbf{D}_k^{(n)}\Big)^{1/2}. \tag{15.25}$$

Clearly, $\mathbf{V}_{\mathbf{j}}^{(n)} \leq \mathbf{P}_n - \sum_{k=1}^N \mathbf{D}_k^{(n)}$. We also have the following result.

Lemma 15.3.8. *There exists an $n_1 \in \mathbb{N}$ such that if $n \geq n_1$ then*

$$\mathrm{tr}[\bar{\sigma}^{(n)}\mathbf{V}_{\vec{j}}^{(n)}] \leq 2^{-n(\chi^*(\Phi)-\frac{2}{3}\epsilon)}, \tag{15.26}$$

and

$$\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{V}_{\vec{j}}^{(n)}]) < 1-\epsilon \tag{15.27}$$

Proof of Lemma 15.3.8. Put $\mathbf{Q}_n = \sum_{k=1}^{N(n)} \mathbf{D}_k^{(n)}$. Note that $\mathbf{Q}_n$ is of the form $\mathbf{Q}_n = \tilde{\mathbf{Q}}_{ml_0}\mathbf{1}_{n-ml_0}$, since $\mathbf{D}_j^{(n)} = \mathcal{D}_j^{ml_0}\mathbf{1}_{n-ml_0}$. Note that $\mathbf{Q}_n$ commutes with $\bar{\Pi}_n$ by condition (i). Now by Lemma 15.3.6, we have

$$\bar{\mathbf{P}}_{ml_0}\bar{\sigma}_{ml_0}\bar{\mathbf{P}}_{ml_0} \leq 2^{-m(H_M(\bar{\phi}_\infty)-\frac{1}{4}\epsilon)}\mathbf{1}^{(ml_0)}$$

and assuming $l_0 \geq L$, we have by Lemma 15.3.4,

$$\begin{aligned}\bar{\mathbf{P}}_{ml_0}\bar{\sigma}_{ml_0}\bar{\mathbf{P}}_{ml_0} &\leq 2^{-m(H(\Phi^{(l_0)}(\bar{\rho}))-\frac{1}{4}(1+\frac{1}{2}l_0)\epsilon)}\mathbf{1}^{(ml_0)} \\ &\leq 2^{-m(\frac{1}{l_0}H(\Phi^{(l_0)}(\bar{\rho}))-\frac{1}{4}\epsilon)}\mathbf{1}^{(ml_0)}\end{aligned}$$

provided that

$$\frac{n-ml_0}{l_0}H(\Phi^{(l_0)}(\bar{\rho})) \leq \frac{1}{4}\left(n-m-\frac{1}{2}(ml_0)\epsilon\right),$$

which holds if $l_0 > 6$ and $m \geq \frac{12}{\epsilon}\log(\dim(\mathbb{H}_B))$, since $\frac{1}{l_0}H(\Phi^{(l_0)}(\bar{\rho})) \leq \log(\dim(\mathbb{H}_B))$.

Using this, we have

$$\begin{aligned}\operatorname{tr}[\bar{\sigma}_n\mathbf{V}_{\vec{j}}^{(n)}] &= \operatorname{tr}[\bar{\sigma}_n(\bar{\Pi}_n-\mathbf{Q}_n)^{1/2}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n(\bar{\Pi}_n-\mathbf{Q}_n)^{1/2}] \\ &= \operatorname{tr}[\bar{\Pi}_n\bar{\sigma}_n\bar{\Pi}_n(\bar{\Pi}_n-\mathbf{Q}_n)^{1/2}\Pi_{\vec{j}}^{(n)}(\bar{\Pi}_n-\mathbf{Q}_n)^{1/2}] \\ &= \operatorname{tr}[\bar{\mathbf{P}}_{ml_0}\bar{\sigma}_{ml_0}\bar{\mathbf{P}}_{ml_0}(\bar{\mathbf{P}}_{ml_0}-\tilde{\mathbf{Q}}_{ml_0})^{1/2}\mathbf{P}_{\vec{j}}^{(ml_0)}(\bar{\mathbf{P}}_{ml_0}-\tilde{\mathbf{Q}}_{ml_0})^{1/2}] \\ &\leq 2^{-n(\frac{1}{l_0}H(\Phi^{(l_0)}(\bar{\rho}_n))-\frac{1}{4}\epsilon)}\operatorname{tr}[(\bar{\mathbf{P}}_{ml_0}-\tilde{\mathbf{Q}}_{ml_0})^{1/2}\mathbf{P}_{\vec{j}}^{(ml_0)}(\bar{\mathbf{P}}_{ml_0}-\tilde{\mathbf{Q}}_{ml_0})^{1/2}] \\ &\leq 2^{-n(\frac{1}{l_0}H(\Phi^{(l_0)}(\bar{\rho}_n)-\frac{1}{4}\epsilon)}\operatorname{tr}[\mathbf{P}_{\vec{j}}^{(ml_0)}].\end{aligned}$$

However,

$$\begin{aligned}\operatorname{tr}[\mathbf{P}_{\vec{j}}^{(ml_0)}] &\leq 2^{m(\bar{H}_M(\{p_j^{(l_0)},\rho_j^{(l_0)}\})+\frac{1}{4}\epsilon)} \\ &\leq 2^{n(\frac{1}{l_0}\bar{H}_M(\{p_j^{(l_0)},\rho_j^{(l_0)}\})+\frac{1}{4}\epsilon)} \\ &\leq 2^{n(\frac{1}{l_0}\sum_j p_j^{(l_0)}H(\rho_j^{(l_0)})+\frac{1}{4}\epsilon)}\end{aligned}$$

where the last inequality follows from the subadditivity of the von Neumann entropy. Inequality (15.26) now follows from (15.12). Since $N(n)$ is maximal, it follows that $\operatorname{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{V}_{\vec{j}}^{(n)}] < 1-\epsilon$. Consequently, (15.27) holds. This proves the lemma. □

Since $N(n)$ is maximal, it follows that for $\mathbf{j} \in \mathbf{W}_n$,

$$\operatorname{tr}[\sigma_{\mathbf{j}}^{(n)}\mathbf{V}_{\mathbf{j}}^{(n)}] \leq 1-2\epsilon. \tag{15.28}$$

We now show that the set $\mathbf{W}_n$ has high probability.

Lemma 15.3.9. *Assume that $\eta > 3\delta$. Then for all $n \geq n_2 = m_1 l_0 \vee m_2 l_0$,*

$$\mathbb{E}(\operatorname{tr}[\sigma_{\vec{j}}^{(n)}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n]) > 1-\eta. \tag{15.29}$$

Proof of Lemma 15.3.9. We write

$$\begin{aligned}&\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n])\\&\quad=\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\Pi_{\vec{j}}^{(n)}])-\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}(\mathbf{1}-\bar{\Pi}_n)\Pi_{\vec{j}}^{(n)}])\\&\qquad-\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}(\mathbf{I}-\bar{\Pi}_n)])\end{aligned}\tag{15.30}$$

The first term equals $\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(ml_0)}\bar{\mathbf{P}}_{\vec{j}}]) > 1-\delta^2$, provided $n \geq n_2$.

Note that

$$\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}(\mathbf{1}-\bar{\Pi}_n)\Pi_{\vec{j}}^{(n)}])=\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(ml_0)}(\mathbf{1}-\bar{P}_{ml_0})\mathbf{P}_{\vec{j}}^{(ml_0)}]),\tag{15.31}$$

and similarly,

$$\begin{aligned}&\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\Pi_{\vec{j}}^{(n)}(\mathbf{1}-\bar{\Pi}_n)])\\&\quad=\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(ml_0)}\mathbf{P}_{\vec{j}}^{(ml_0)}(\mathbf{1}-\bar{\mathbf{P}}_{ml_0})]).\end{aligned}\tag{15.32}$$

Using (15.31) and (15.32), the last two terms of the right-hand side of (15.30) can be bounded using the Cauchy–Schwarz inequality and Lemma 15.3.6 as follows:

$$\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}(\mathbf{1}-\bar{\Pi}_n)\Pi_{\vec{j}}^{(n)}])\leq\delta$$

and

$$\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\Pi_{\vec{j}}^{(n)}(\mathbf{1}-\bar{\Pi}_n)])\leq\delta$$

provided $n \geq n_1$. Choosing $n_3 = n_1 \vee n_2$ and $\delta^2 + 2\delta < \eta$ the result follows. This proves the lemma. □

Lemma 15.3.10. *Assume $\eta < \frac{1}{3}\epsilon$ and $\eta > 3\delta$. Then for $n \geq n_3 = n_1 \vee n_2$,*

$$\mathrm{tr}\left[\bar{\sigma}_n\sum_{k=1}^{N}\mathbf{D}_k^{(n)}\right]=\mathbb{E}\left(\mathrm{tr}\left[\sigma_{\vec{j}}^{(n)}\sum_{k=1}^{N}\mathbf{D}_k^{(n)}\right]\right)\geq\eta^2.\tag{15.33}$$

Proof of Lemma 15.3.10. Define

$$\mathbf{Q}_n'=\bar{\Pi}_n-(\bar{\Pi}_n-\mathbf{Q}_n)^{1/2},\tag{15.34}$$

where $\mathbf{Q}_n = \sum_{k=1}^N \mathbf{D}_k^{(n)}$. It follows that

$$\begin{aligned}1-\epsilon&\geq\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}(\bar{\Pi}_n-\mathbf{Q}_n')\Pi_{\vec{j}}^{(n)}(\bar{\Pi}_n-\mathbf{Q}_n')]\}\\&=\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n]\}-\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n]\}\\&\quad-\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\Pi_n\Pi_{\vec{j}}^{(n)}\mathbf{Q}_n']\}+\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'\Pi_{\vec{j}}^{(n)}\mathbf{Q}_n']\}\end{aligned}\tag{15.35}$$

Since the last term is positive, by Lemma 15.3.9 we have

$$\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n] + \mathrm{tr}[\sigma_{\vec{j}}^{(n)}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}\mathbf{Q}_n']\} > \epsilon - \eta > 2\eta. \tag{15.36}$$

On the other hand, using the Cauchy–Schwarz inequality (1.2) for each term, we have

$$\begin{aligned}
&\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n] + \mathrm{tr}[\sigma_{\vec{j}}^{(n)}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}\mathbf{Q}_n']\} \\
&\quad \le 2\{\mathbb{E}(\mathrm{tr}[\mathbf{Q}_n'\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'])\}^{1/2}\{\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\bar{\Pi}_n\Pi_{\vec{j}}^{(n)}\bar{\Pi}_n])\}^{1/2} \\
&\quad \le 2\{\mathbb{E}(\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'^2])\}^{1/2}.
\end{aligned} \tag{15.37}$$

Thus,

$$\mathbb{E}\{\mathrm{tr}[\sigma_{\vec{j}}^{(n)}\mathbf{Q}_n'^2]\} \ge \eta^2 \tag{15.38}$$

To complete the proof of this lemma, we now claim that

$$\mathbf{Q}_n \ge \mathbf{Q}_n'^2. \tag{15.39}$$

Indeed, this follows on the domain of $\bar{\Pi}_n$ from the inequality $1-(1-x)^2 \ge x^2$ for $0 \le x \le 1$ This proves the lemma. □

To complete the proof of Theorem 13.2.5, we now have by assumption

$$\mathrm{tr}[\bar{\sigma}_n\mathbf{D}_k^{(n)}] \le 2^{-n(H(\bar{\sigma})-\bar{H}-\frac{2}{3}\epsilon)} \tag{15.40}$$

for all $k = 1, \ldots, N(n)$. On the other hand, choosing $\eta < \frac{1}{3}\epsilon$ and $\delta < \frac{1}{3}\eta$, we have by Lemma 13.2.9,

$$\mathrm{tr}\left[\bar{\sigma}_n\sum_{k=1}^{N}\mathbf{D}_k^{(n)}\right] \ge \eta^2,$$

provided that $n \ge n_3$. It follows from the definition of $N(n)$ that

$$N(n) \ge \eta^2 2^{n(H(\bar{\sigma})-\bar{H}-\frac{2}{3}\epsilon)} \ge 2^{n(H(\bar{\sigma})-\bar{H}-\epsilon)}$$

for $n \ge n_3$ and $n \ge -\frac{6}{\epsilon}\log\eta$. This proves the theorem. □

To complete the proof of Proposition 15.3.3, we now have by assumption,

$$\mathrm{tr}[\bar{\sigma}^{(n)}\mathbf{D}_k^{(n)}] \le 2^{-n(\chi^*(\Phi)-\frac{2\epsilon}{3})}, \quad \forall k = 1, 2, \ldots, N(n).$$

On the other hand, choosing $\eta < \frac{1}{3}\epsilon$ and $\delta < \frac{1}{3}\eta$, we have by Lemma 15.3.10,

$$\operatorname{tr}\left[\bar{\sigma}\sum_{k=1}^{N}\mathbf{D}_k\right] \geq \eta^2,$$

provided $n \geq n_3$. It follows that

$$N(n) \geq \eta^2 2^{n(\chi^*(\Phi)-\frac{2}{3}\epsilon)} \geq 2^{n(\chi^*(\Phi)-\epsilon)}$$

for all $n \geq n_3$ and $n \geq -\frac{6}{\epsilon}\log\eta$. This proves the proposition. □

15.4 Channels with general Markov memory

We consider the quantum channel with Markovian memory $\Phi^{(\infty)} = \{\Phi^{(n)}\}_{n=1}^{+\infty}$ defined in (15.4), where $\{X_n\}_{n=1}^{+\infty}$ is a general Markov chain defined on a probability space $(\Omega, \Sigma, \Pr)$ with finite states $\mathbb{I}$.

In this section, we first construct a preamble for channels with general Markovian memory and then explore the formula for computing its classical capacity (see Datta and Dorlas [33]).

15.4.1 Preamble for Markovian memory

Let $\mathcal{C}$ be the set of communicating classes of the Markov chain $\{X_n\}_{n=0}^{+\infty}$ and let $C \in \mathcal{C}$ be a certain communicating class for which

$$\gamma_C = \sum_{i\in C}\gamma_i > 0,$$

where $\{\gamma_i\}_{i\in\mathbb{I}}$ denotes the invariant/stationary distribution of the Markov chain. We disregard all communicating classes C, where $\gamma_C = 0$. For $C \in \mathcal{C}$ with $\gamma_C > 0$, define

$$\Phi_C^{(n)}(\rho_{\Lambda_n,A}^{(n)}) := \frac{1}{\gamma_C}\sum_{i_1,\ldots,i_n\in C}\gamma_{i_1}q_{i_1i_2}\cdots q_{i_{n-1}i_n}(\Phi_{i_1}\otimes\cdots\otimes\Phi_{i_n})(\rho_{\Lambda_n,A}^{(n)}), \tag{15.41}$$

which represents the restriction of the Markovian memory of the channel to the class C. Notice that the Markov chain restricted to $C \in \mathcal{C}$ is necessarily irreducible, and is either aperiodic or periodic with a single period. In fact,

$$\mathcal{C} = \mathcal{C}_{\text{aper}} \cup \mathcal{C}_{\text{per}},$$

where $\mathcal{C}_{\text{aper}}$ denotes the set of communicating classes in $\mathcal{C}$, which are aperiodic, while $\mathcal{C}_{\text{per}}$ denotes the set of communicating classes in $\mathcal{C}$, which are periodic.

In the following, we try to distinguish the $\Phi_C^{(n)}$ of the channel Φ for different classes of communication class C. To do so, we add a preamble of quantum states to the input state encoding each message in the set $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{M_n}\}$.

Let us first sketch the idea behind adding such a preamble. Helstrom [65] showed that two quantum states σ_1 and σ_2, occurring with *a priori* probabilities p_1 and p_2, respectively, can be distinguished with an asymptotically vanishing probability of error, if a suitable collective measurement is performed on its m-fold tensor products $\sigma_1^{\otimes m}$ and $\sigma_2^{\otimes m}$, for a large enough $m \in \mathbb{N}$. The relevant projection operators, which we denote by Π_+ and Π_-, are the orthogonal projections onto the positive and negative eigenspaces of the difference operator $\mathbf{A}_m = p_1\sigma_1^{\otimes m} - p_2\sigma_2^{\otimes m}$ in Helstrom's scheme.

We generalize the Helstrom result to distinguish between the different classes $\Phi_C^{(n)}$, where the preamble is given by an m-fold tensor product of suitable states described as follows. If the preamble is given by a state $\omega^{\otimes m}$, then by using Helstrom's result, we can construct a POVM, which distinguishes between the output states $\sigma_C^{(n)} := \Phi_C^{(n)}(\omega^{\otimes m})$ corresponding to the different classes $\Phi_C^{(n)}$. The outcome of this POVM measurement would in turn serve to determine which class of the channel is being used for transmission.

1. Following the idea sketched above (see Datta and Dorlas [33]), we first show that there exists a preamble that can distinguish between the channels for different communicating classes. In fact, we can do more than just distinguish the channel for different communication classes. In the case of periodic classes, we also want to distinguish between initial states of the class.

   We, therefore, subdivide the problem into the following four possibilities:
   (A) to distinguish between two aperiodic classes;
   (B) to distinguish between an aperiodic class C_1 and an initial state σ_2 of a periodic class C_2;
   (C) to distinguish between two periodic classes C_1 and C_2 and
   (D) to distinguish between the states of a single periodic class.

The existence of preamble for each of the above four cases will be provided in part (A) to part (D) below.

(A) The following shows how we can distinguish between two aperiodic classes C_1 and C_2. We can obviously assume that the $\Phi_{C_1}^{(n)} \neq \Phi_{C_2}^{(n)}$ for some n: otherwise the classes are identical and we can combine their probabilities into one aperiodic class. This means that for any pair of aperiodic classes C_1, C_2 there exists $n = n(C_1, C_2)$ and a state $\omega^{(n)} = \omega_{C_1,C_2}^{(n)}$ such that $\Phi_{C_1}^{(n)}(\omega^{(n)}) \neq \Phi_{C_2}^{(n)}(\omega^{(n)})$. In fact, in most cases we can take $n = 1$, and we shall assume this for simplicity in the following, even though this is not necessary.

Recall from (6.1) that the fidelity of two states σ and ρ is defined by

$$F(\sigma,\rho) = \mathrm{tr}[\sqrt{\sigma^{1/2}\rho\sigma^{1/2}}],$$

which can be used to measure the distinguishability of σ and ρ. Since $\Phi_{C_1}^{(1)}(\omega) \neq \Phi_{C_2}^{(1)}(\omega)$, we then have

$$F(\Phi_{C_1}^{(1)}(\omega), \Phi_{C_2}^{(1)}(\omega)) \le f < 1, \tag{15.42}$$

for all distinct pairs C_1, C_2 with $C_1 < C_2$ in some arbitrary ordering of the set of aperiodic classes, $\mathcal{C}_{\text{aper}}$.

The following lemma shows that the distinguishability of any two aperiodic classes C_1 and C_2 asymptotically vanishes via the use of quantum fidelity.

Lemma 15.4.1. *For any two aperiodic classes C_1 and C_2,* (15.42) *implies that*

$$\lim_{m\to+\infty} F(\Phi_{C_1}^{(m)}(\omega_{C_1,C_2}^{\otimes m}), \Phi_{C_2}^{(m)}(\omega_{C_1,C_2}^{\otimes m})) = 0. \tag{15.43}$$

Proof. Since $f < 1$, we can choose $\alpha > 0$ so small that $1+\alpha < f^{-1}$. First, let k be so large that

$$(1-\alpha)\gamma_j < \sum_{i_2,\dots,i_{k-1}} q_{ii_2}\cdots q_{i_{k-1}j} < (1+\alpha)\gamma_j \tag{15.44}$$

for all $i, j \in \mathbb{I}$. By Theorem 6.1.10, we can choose a POVM $\{\mathbf{E}_r\}_r$ such that

$$F(\sigma_1, \sigma_2) = \sum_r \sqrt{\operatorname{tr}[\sigma_1 \mathbf{E}_r]\operatorname{tr}[\sigma_2 \mathbf{E}_r]}, \tag{15.45}$$

where we denote

$$\sigma_1 = \Phi_{C_1}^{(1)}(\omega_{C_1,C_2}) \quad \text{and} \quad \sigma_2 = \Phi_{C_2}^{(1)}(\omega_{C_1,C_2}). \tag{15.46}$$

Then we have

$$\begin{aligned} &F(\Phi_{C_1}^{(mk+m)}(\omega_{C_1,C_2}^{\otimes(mk+m)}), \Phi_{C_2}^{(mk+m)}(\omega_{C_1,C_2}^{\otimes(mk+m)})) \\ &\quad \le \sum_{r_1,r_2,\dots,r_m} (\operatorname{tr}[\Phi_{C_1}^{(mk+m)}(\omega_{C_1,C_2}^{\otimes(mk+m)}) \otimes_{i=1}^m (\mathbf{E}_{r_i} \otimes \mathbf{1}_k)] \\ &\qquad \times \operatorname{tr}[\Phi_{C_2}^{(mk+m)}(\omega_{C_1,C_2}^{\otimes(mk+m)}) \otimes_{i=1}^m (\mathbf{E}_{r_i} \otimes \mathbf{1}_k)])^{1/2} \\ &\quad \le (1+\alpha)^{m-1} \prod_{i=1}^m \Big(\sum_{r_i} \sqrt{\operatorname{tr}[\sigma_1 \mathbf{E}_{r_i}]\operatorname{tr}[\sigma_2 \mathbf{E}_{r_i}]} \Big) \\ &\quad = (1+\alpha)^{m-1} F(\sigma_1, \sigma_2)^m \to 0 \quad \text{as } m \to +\infty. \end{aligned} \tag{15.47}$$

This proves the lemma. □

(B) Next, consider the second case, i. e., to distinguish an aperiodic class C_1 and an initial state j of a periodic class C_2.

In this case, there exists a state $\omega = \omega_{C_1,j}$ on $\mathbb{H}_A$ such that

$$f := F(\Phi_{C_1}^{(1)}(\omega_{C_1,j}), \Phi_j^{(1)}(\omega_{C_1,j})) < 1. \tag{15.48}$$

The following lemma shows that the channels Φ_{C_1} and Φ_j are asymptotically nondistinguishable.

Lemma 15.4.2. *Let C_1 be an aperiodic and C_2 be a periodic class with period $L = L(C_2)$. Let $j \in C_2$ and choose $\omega_{C_1,j}$ as above. Then*

$$\lim_{m\to+\infty} F(\Phi_{C_1}^{(m)}(\omega_{C_1,j}^{(m)}), \Phi_j^{(m)}(\omega_{C_1,j}^{(m)})) = 0. \tag{15.49}$$

Proof. We proceed as in Lemma 15.4.1 and choose $\alpha > 0$ so small that $1 + \alpha < f^{-1}$ and let k be so large that (15.44) holds and in addition such that k is a multiple of L. Again, we let $\{\mathbf{E}_r\}_r$ be a POVM such that

$$F(\sigma_1, \sigma_2) = \sum_r \sqrt{\operatorname{tr}[\sigma_1 \mathbf{E}_r] \operatorname{tr}[\sigma_2 \mathbf{E}_r]}, \tag{15.50}$$

where now

$$\sigma_1 = \Phi_{C_1}^{(1)}(\omega_{C_1,j}) \quad \text{and} \quad \sigma_2 = \Phi_j^{(1)}(\omega_{C_1,j}). \tag{15.51}$$

Then we have

$$\begin{aligned}
&F(\Phi_{C_1}^{(mk+m)}(\omega_{C_1,j}^{\otimes(mk+m)}), \Phi_{C_2,j}^{(mk+m)}(\omega_{C_1,j}^{\otimes(mk+m)})) \\
&\le \sum_{r_1,r_2,\dots,r_m} \Bigg(\operatorname{tr}[\Phi_{C_1}^{(mk+m)}(\omega_{C_1,j}^{\otimes(mk+m)}) \otimes_{i=1}^m (\mathbf{E}_{r_i} \otimes \mathbf{1}_k)] \\
&\quad \times \prod_{l=1}^m \operatorname{tr}[\Phi_{C_2,j}^{(mk+m)}(\omega_{C_1,j}^{\otimes(mk+m)}) \mathbf{E}_{r_l}] \Bigg)^{1/2} \\
&\le \sum_{r_1,r_2,\dots,r_m} (1+\alpha)^{\frac{m-1}{2}} \prod_{i=1}^m \sqrt{\operatorname{tr}[\sigma_1 \mathbf{E}_{r_i}] \operatorname{tr}[\sigma_2 \mathbf{E}_{r_i}]} \\
&= (1+\alpha)^{\frac{m-1}{2}} F(\sigma_1, \sigma_2)^m \to 0 \quad \text{as } m \to +\infty.
\end{aligned} \tag{15.52}$$

This proves the lemma. □

(C) Distinguishability of two periodic classes vanishes asymptotically as illustrated below.

Lemma 15.4.3. *If C_1 and C_2 are two different periodic classes with periods $L(C_1)$ and $L(C_2)$, respectively, then there exists a state $\omega_{C_1,C_2}^{(L)}$ on $\mathbb{H}_A^{\otimes L}$, where $L = L(C_1)L(C_2)$ such that*

$$\lim_{m\to+\infty} F(\Phi_{C_1}^{(mL)}((\omega_{C_1,C_2}^{(L)})^{\otimes m}), \Phi_{C_2}^{(mL)}((\omega_{C_1,C_2}^{(L)})^{\otimes m})) = 0. \tag{15.53}$$

Proof. Since the two periodic classes are distinct, there exists a state $\omega_{C_1,C_2}^{(L)}$ such that

$$\Phi_{C_1}^{(L)}(\omega_{C_1,C_2}^{(L)}) \neq \Phi_{C_2}^{(L)}(\omega_{C_1,C_2}^{(L)}). \tag{15.54}$$

(In fact, we can take L to be the least common multiple of $L(C_1)$ and $L(C_2)$.) Then writing $\omega = \omega_{C_1,C_2} \otimes \varphi^{\otimes k}$, where φ is an arbitrary state on $\mathbb{H}_A$ and k is so large that (15.44) holds, and we have

$$\begin{aligned}&\lim_{m\to+\infty} F(\Phi_{C_1}^{(mL+mk)}(\omega^{\otimes m}), \Phi_{C_2}^{(mL+mk)}(\omega^{\otimes m}))\\ &\quad\le \lim_{m\to+\infty} (1+\alpha)^m F((\Phi_{C_1}^{(L)}(\omega))^{\otimes m}, (\Phi_{C_2}^{(L)}(\omega))^{\otimes m}) = 0.\end{aligned} \tag{15.55}$$

This proves the lemma. □

(D) Finally, to distinguish the initial states of a given periodic class C, notice first of all that the corresponding CPTP maps Φ_i need not all be distinct. However, we may assume that there is no internal periodicity of these maps within a periodic class; otherwise the class can be contracted to a single such period. This means, that for any two states $i, j \in C$ there exists $l \le L(C) - 1$ such that $\Phi_{i+l} \neq \Phi_{j+l}$. Then choose $\omega = \omega_{i,j}$ such that

$$f := F(\Phi_{i+l}(\omega), \Phi_{j+l}(\omega)) < 1. \tag{15.56}$$

Lemma 15.4.4. *If C is a periodic class with period $L(C)$, $i, j \in C$ and ω is a state as above, then*

$$\lim_{m\to+\infty} F(\Phi_i^{(m)}(\omega^{\otimes m}), \Phi_j^{(m)}(\omega^{\otimes m})) = 0. \tag{15.57}$$

Proof. We have the following inequality:

$$\begin{aligned}&F(\Phi_i^{(m)}(\omega^{\otimes m}), \Phi_j^{(m)}(\omega^{\otimes m}))\\ &\quad= (F(\Phi_i^{(L)}(\omega^{\otimes L}), \Phi_j^{(L)}(\omega^{\otimes L})))^m\\ &\quad\le (F(\Phi_{i+l}(\omega), \Phi_{j+l}(\omega)))^m = f^m \to 0 \quad \text{as } m\to+\infty.\end{aligned} \tag{15.58}$$

This proves the lemma. □

2. We now introduce, in each of the four cases, difference operators $\mathbf{A}_{C_1,C_2}^{(m)}$, $\mathbf{A}_{C_1,j}^{(m)}$ and $\mathbf{A}_{i,j}^{(m)}$ with i, j in a periodic class, and corresponding projections $\Pi_{C_1,C_2}^{\pm}$, $\Pi_{C_1,j}^{\pm}$ and $\Pi_{i,j}^{\pm}$ onto their positive and negative eigenspaces, which serve to distinguish the different possibilities, as in Datta and Dorlas [33]. The difference operators are defined by

$$\mathbf{A}^{(m)}_{C_1,C_2} := \gamma_{C_1}\big(\Phi^{(m)}_{C_1}(\omega_{C_1,C_2})\big)^{\otimes m} - \gamma_{C_2}\big(\Phi^{(m)}_{j}(\omega_{C_1,C_2})\big)^{\otimes m}; \tag{15.59}$$

$$\mathbf{A}^{(m)}_{C_1,j} := \gamma_{C_1}\big(\Phi^{(m)}_{C_1}(\omega_{C_1,j})\big)^{\otimes m} - \gamma_{j}\big(\Phi^{(m)}_{C_2}(\omega_{C_1,j})\big)^{\otimes m}; \tag{15.60}$$

and

$$\mathbf{A}^{(m)}_{i,j} := \gamma_{i}\big(\Phi^{(m)}_{i}(\omega_{i,j})\big)^{\otimes m} - \gamma_{j}\big(\Phi^{(m)}_{j}(\omega_{i,j})\big)^{\otimes m}. \tag{15.61}$$

The following lemma is due originally to Datta and Dorlas [33].

Lemma 15.4.5. *Suppose that for a given $\delta > 0$,*

$$\big|\mathrm{tr}[|\mathbf{A}^{(m)}_{C_1,C_2}|] - (\gamma_{C_1} + \gamma_{C_2})\big| \leq \delta. \tag{15.62}$$

Then

$$\big|\mathrm{tr}[\Pi^{+}_{c,c'}(\Phi^{(m)}_{c}(\omega^{\otimes m}_{c,c'}))] - 1\big| \leq \frac{\delta}{2\gamma_c} \tag{15.63}$$

and

$$\big|\mathrm{tr}[\Pi^{-}_{c,c'}(\Phi^{(m)}_{c}(\omega^{\otimes m}_{c,c'}))] - 1\big| \leq \frac{\delta}{2\gamma_{c'}}. \tag{15.64}$$

Here, c, c' denote either two different classes C_1, C_2 or one aperiodic class C and an initial state j in a periodic class, or two different initial states in the same periodic class. To compare the outputs of all the different branches of the channel, we define projections $\tilde{\Pi}$ on the tensor product space $\mathbb{H}_B^{\otimes mM}$, where

$$M = M_1 + M_2 + M_3 + M_4, \tag{15.65}$$

with:
(i) M_1 is the total number of pairs of aperiodic classes;
(ii) M_2 is the total number of pairs of periodic classes;
(iii) M_3 is the total number of pairs of aperiodic classes and initial states of periodic classes; and
(iv) M_4 is the total number of pairs of states in the same periodic class.

We introduce an arbitrary order on the classes $C \in \mathcal{C}$ assuming $C_1 < C_2$ if $C_1 \in \mathcal{C}_{\mathrm{aper}}$ and $C_2 \in \mathcal{C}_{\mathrm{per}}$. Then we put

$$\tilde{\Pi}_c = \bigotimes_{\{c',c''\}:c'<c''} \Gamma^{(c)}_{c',c''}, \tag{15.66}$$

where

$$\Gamma^{(c)}_{c',c''} = \begin{cases} \mathbf{I}_m & \text{if } c' \neq c \text{ and } c'' \neq c \\ \Pi^-_{c',c} & \text{if } c'' = c \\ \Pi^+_{c,c''} & \text{if } c' = c. \end{cases}$$

If follows from the fact that $\Pi^+_{c',c''}\Pi^-_{c',c''} = \mathbf{0}$ the projections $\tilde{\Pi}_c$ are also disjoint:

$$\tilde{\Pi}_{c_1}\tilde{\Pi}_{c_2} = \mathbf{0} \quad \text{for } c_1 \neq c_2. \tag{15.67}$$

We use the following lemma.

Lemma 15.4.6. *For all aperiodic classes C,*

$$\lim_{m\to+\infty} \operatorname{tr}[\tilde{\Pi}_C \Phi^{(mM)}_C(\omega^{(mM)})] = 1, \tag{15.68}$$

and for all periodic classes C and all states $i \in C$,

$$\lim_{m\to+\infty} \operatorname{tr}[\tilde{\Pi}_{C,i} \Phi^{(mM)}_{C,i}(\omega^{(mM)})] = 1. \tag{15.69}$$

Proof. Notice that for all (c, c'),

$$\begin{aligned} &\lim_{m\to+\infty} F\big(\gamma_c \Phi^{(mM)}_c(\omega^{\otimes m}_{c,c'}), \gamma_{c'} \Phi^{(mM)}_{c'}(\omega^{\otimes m}_{c,c'})\big) \\ &= \lim_{m\to+\infty} \sqrt{\gamma_c\gamma_{c'}} F\big(\Phi^{(mM)}_c(\omega^{\otimes m}_{c,c'}), \Phi^{(mM)}_{c'}(\omega^{\otimes m}_{c,c'})\big) \\ &= 0 \end{aligned} \tag{15.70}$$

Using the inequalities,

$$\operatorname{tr}[\mathbf{A}_1] + \operatorname{tr}[\mathbf{A}_2] - 2F(\mathbf{A}_1, \mathbf{A}_2) \leq \|\mathbf{A}_1 - \mathbf{A}_2\|_1 \leq \operatorname{tr}[\mathbf{A}_1] + \operatorname{tr}[\mathbf{A}_2] \tag{15.71}$$

for any two positive operators $\mathbf{A}_1$ and $\mathbf{A}_2$, we have

$$\lim_{m\to+\infty} \big|\operatorname{tr}[|\mathbf{A}^{(m)}_{c,c'}|] - (\gamma_i + \gamma_j)\big| \leq \lim_{m\to+\infty} \delta_m = 0, \tag{15.72}$$

since

$$\operatorname{tr}[|\mathbf{A}^{(m)}_{c,c'}|] = \big|F\big(\gamma_c \Phi^{(mM)}_c(\omega^{\otimes m}_{c,c'}), \gamma_{c'} \Phi^{(mM)}_{c'}(\omega^{\otimes m}_{c,c'})\big)\big|. \tag{15.73}$$

We now replace m by $m' = m + k$, where $k \in \mathbb{N}$ is large enough so that (15.44) holds, and define

$$\omega^{(m'M)} := \bigotimes_{c_1,c_2} \omega^{\otimes m+k}_{c_1,c_2}. \tag{15.74}$$

Using (15.44) to separate the different classes, we then have for any $C \in \mathcal{C}_{\text{aper}}$,

$$\begin{aligned}1 &\geq \operatorname{tr}\Big[\tilde{\Pi}_C \Phi_C^{(m'M)}\Big(\bigotimes_{C_1<C_2} \omega_{C_1,C_2}^{\otimes(m+k)}\Big)\Big] \\ &\geq (1-\alpha)^M \prod_{C'\in\mathcal{C}_{\text{aper}}:C'<C} \operatorname{tr}[\Pi_{C',C}^{-}\Phi_C^{(m)}(\omega_{C',C}^{\otimes m})] \\ &\quad\times \prod_{C''\in\mathcal{C}_{\text{aper}}:C''>C} \operatorname{tr}[\Pi_{C,C''}^{+}\Phi_C^{(m)}(\omega_{C,C''}^{\otimes m})] \\ &\quad\times \prod_{C'\in\mathcal{C}_{\text{aper}}:C'<C}\ \prod_{i'\in C'} \operatorname{tr}[\Pi_{C',i'}^{+}\Phi_C^{(m)}(\omega_{C',i'}^{\otimes m})] \\ &\geq (1+\alpha)^M\Big(1-\frac{\delta_m}{2\gamma_C}\Big)^{|\mathcal{C}_{\text{aper}}|-1+\sum_{C'\in\mathcal{C}_{\text{apper}}}|C'|} \to 1 \end{aligned} \tag{15.75}$$

since $\delta_m \to 0$ as $m \to +\infty$. The last inequality follows from Lemma 15.4.5. The analogous result, (15.69), for periodic classes, is proved in a similar manner. This proves the lemma. □

15.4.2 Classical capacity

This section investigates classical capacity of channels with general Markovian memory.

Due to the results in Section 15.4.1, we only need to consider the classical capacity of the channel restricted to a single communication class $C \in \mathcal{C}$, since distinction of channel between different classes and between different states in a periodic class vanishes asymptotically as the number of use increases to infinity.

Let $\mathcal{C}$ be the collection of communicating classes of the Markov chain $\{X_n\}_{n=0}^{+\infty}$ as discussed earlier.

We consider the following two cases: (I) $C \in \mathcal{C}_{\text{aper}}$; and (II) $C \in \mathcal{C}_{\text{per}}$ as follows.

(I) If $C \in \mathcal{C}_{\text{aper}}$, we define, for any ensemble $\{p_j^{(n)}, \rho_j^{(n)}\}$ of states on $\mathbb{H}_A^{\otimes n}$, the mean Holevo quantity for the class C as

$$\bar{\chi}_C^{(n)}(\{p^{(n)},\rho^{(n)}\}) = \frac{1}{n}H\Big(\sum_j p_j^{(n)}\Phi_C^{(n)}(\rho_j^{(n)})\Big) - \frac{1}{n}\sum_j p_j^{(n)} H(\Phi_C^{(n)}(\rho_j^{(n)})), \tag{15.76}$$

where $H(\cdot) : \mathcal{S}(\mathbb{H}_A) \to [0,+\infty]$ is the Shannon entropy function.

(II) If $C \in \mathcal{C}_{\text{per}}$ is periodic, with period L, then $C = \{i_0, i_1, \dots, i_{L-1}\}$ for certain $i_0, \dots, i_{L-1} \in \mathbb{I}$, and $q_{i_k i_{k+1}} = 1$ for $k = 0, \dots, L-2$ and $q_{i_{L-1}, i_0} = 1$. In this case,

$$\gamma_i = \frac{1}{L}\gamma_C \quad (i \in C) \tag{15.77}$$

and we set

$$\bar{\chi}_C^{(n)}(\{p^{(n)}, \rho^{(n)}\}) = \frac{1}{nL}\sum_{i\in C}\chi_{C,i}^{(n)}(\{p_j^{(n)}, \rho_j^{(n)}\}), \tag{15.78}$$

where, for $k \in \{1, 2, \dots, L-1\}$,

$$\chi_{C,i_k}^{(n)}(\{p_j^{(n)}, \rho_j^{(n)}\}) = H\Big((\Phi_{i_k} \otimes \Phi_{i_{k+1}} \otimes \cdots \otimes \Phi_{i_{k+n-1}})\Big(\sum_j p_j^{(n)}\rho_j^{(n)}\Big)\Big) \\ - \sum_j p_j^{(n)} H((\Phi_{i_k} \otimes \Phi_{i_{k+1}} \otimes \cdots \otimes \Phi_{i_{k+n-1}})(\rho_j^{(n)})), \tag{15.79}$$

and the indices in the subscripts being taken modulo L.

In either of cases (I) and (II), the average probability of error for the code $\mathfrak{C}^{(n)}$ is given by

$$\bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)}) := \frac{1}{M_n}\sum_{i=1}^{M_n}(1 - \mathrm{tr}[\Phi^{(n)}(\rho_i^{(n)})\mathbf{D}_i^{(n)}]). \tag{15.80}$$

If there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, there exists a sequence of codes $\{\mathfrak{C}^{(n)}\}_{n=1}^{+\infty}$, of sizes $M_n \geq 2^{nR}$, for which $\bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)}) \to 0$ as $n \to \infty$, then R is said to be an *achievable rate*.

The following result is due to Datta and Dorlas [33].

Theorem 15.4.7 (Datta and Dorla [33]). *The classical capacity of a quantum channel with general Markovian memory, $\Phi^{(\infty)} = \{\Phi^{(n)}\}_{n=1}^{+\infty}$, defined by* (15.4) *is given by*

$$C^{\mathrm{mar}}(\Phi) = \lim_{n\to+\infty}\sup_{\{p_j^{(n)},\rho_j^{(n)}\}}\Big[\min_{C\in\mathcal{C}}\bar{\chi}_C^{(n)}(\{p_j^{(n)}, \rho_j^{(n)}\})\Big], \tag{15.81}$$

where $\bar{\chi}_C^{(n)}(\{p^{(n)}, \rho^{(n)}\})$ is given by (15.76) *for $C \in \mathcal{C}_{\mathrm{aper}}$ and given by* (15.78) *for $C \in \mathcal{C}_{\mathrm{per}}$.*

We postpone the proof of the above theorem until the following results are established.

First, we prove that the following lemma shows that the limit in (15.81) exists.

Lemma 15.4.8. *Let $\Phi^{(\infty)} = \{\Phi^{(n)}\}_{n=1}^{+\infty}$ be a quantum channel with general Markov memory defined by* (15.4). *Then the following limit exist:*

$$\lim_{n\to+\infty}\sup_{\{p_j^{(n)},\rho_j^{(n)}\}}\Big[\min_{C\in\mathcal{C}}\bar{\chi}_C^{(n)}(\{p_j^{(n)}, \rho_j^{(n)}\})\Big], \tag{15.82}$$

where $\bar{\chi}_C^{(n)}(\{p^{(n)}, \rho^{(n)}\})$ is given by (15.76) *for $C \in \mathcal{C}_{\mathrm{aper}}$ and given by* (15.78) *for $C \in \mathcal{C}_{\mathrm{per}}$.*

Proof. For each $n \in \mathbb{N}$, write

$$\bar{\chi}_n = \sup_{\{p_j^{(n)},\rho_j^{(n)}\}} \Big[\min_{C\in\mathcal{C}} \bar{\chi}_C^{(n)}(\{p_j^{(n)},\rho_j^{(n)}\})\Big]. \tag{15.83}$$

We first assume that for any $\delta > 0$ there exist n_0 and m_0 such that for all $n' \geq n_0$ and $n \geq m_0 n'$, $\bar{\chi}_n \geq \bar{\chi}_{n'} - \delta$. Under this assumption, the lemma follows because obviously, $0 \leq \bar{\chi}^n \leq \log(\dim(\mathbb{H}_B))$, and it follows that

$$\liminf_{n\to+\infty} \bar{\chi}_n \geq \bar{\chi}_{n\prime} - \delta$$

and hence, $\liminf_{n\to+\infty} \bar{\chi}_n \geq \limsup_{n'\to+\infty} \bar{\chi}_{n'} - \delta$, where $\delta > 0$ is arbitrary. Therefore, $\lim_{n\to+\infty} \bar{\chi}_n$ exists.

We now prove that for any $\delta > 0$ there exist n_0 and m_0 such that for all $n' \geq n_0$ and $n \geq m_0 n'$, $\bar{\chi}_n \geq \bar{\chi}_{n'} - \delta$ as follows. Let n' be large, and suppose that $\{p^{(n')},\rho^{(n')}\}$ is a maximizing ensemble for (15.83), with n replaced by n', i. e.,

$$\bar{\chi}_{n'} = \min_{C\in\mathcal{C}} \bar{\chi}_C^{(n')}(\{p_j^{(n')},\rho_j^{(n')}\}).$$

Given $n \geq n'$, put $m = [n/n']$ (the integral part of n/n') and $l = n - mn'$. Define the states $\rho_{\vec{j}}^{(n)} = \bigotimes_{r=1}^m \rho_{j_r}^{(n)} \otimes \rho_{j_{m+1}}^{(l)}$, $\vec{j} = (j_1, j_2, \ldots, j_{m+1})$, where $\rho^{(l)}$ is the reduced state on $\mathbb{H}_A^{\otimes n}$. Then $\bar{\rho}^{(n)} = \bigotimes_{r=1}^m \bar{\rho}^{(n)} \otimes \bar{\rho}^{(l)}$, with $\bar{\rho}^{(n')} := \sum_j p_j^{(n')}\rho_j^{(n')}$. We now write for any class $C \in \mathcal{C}$,

$$\begin{aligned}\Phi_C^{(n)}(\bar{\rho}^{(n)}) = \sum_{i_1,\ldots,i_{m+1}\in C}\ \sum_{i'_1,\ldots,i'_{m+1}\in C} &\frac{q_{i'_1 i_2}}{\gamma_{i_2}} \cdots \frac{q_{i'_m i_{m+1}}}{\gamma_{i_{m+1}}} \\ &\times \sigma_C^{(n')}(i_1,i'_1)\otimes\cdots\otimes\sigma_C^{(n')}(i_m,i'_m)\otimes\sigma_C^{(l)}(i_{m+1},i'_{m+1}),\end{aligned} \tag{15.84}$$

where

$$\sigma_C^{(n')}(i,i') = \sum_{i_2,\ldots,i_{n'-1}} \gamma_i q_{ii_2} q_{i_2 i_3} \cdots q_{i_{n'-1}i'} (\Phi_i \otimes \Phi_{i_2} \otimes \cdots \otimes \Phi_{i'})(\bar{\rho}^{(n')}) \tag{15.85}$$

and similarly for $\sigma_C^{(l)}(i,i')$. Let $\gamma = \bigwedge_{i\in\mathbb{I}} \gamma_i := \min_{i\in\mathbb{I}} \gamma_i$. Using the positivity of the density operators and the fact that $q_{ij} \leq 1 \leq \gamma_i/\gamma$, we obtain a simple operator inequality

$$\Phi_C^{(n)}(\bar{\rho}^{(n)}) \leq \frac{1}{\gamma^m} \underbrace{\Phi_C^{(n')}(\bar{\rho}^{(n')})\otimes\Phi_C^{(n')}(\bar{\rho}^{(n')})\otimes\cdots\otimes\Phi_C^{(n')}(\bar{\rho}^{(n')})}_{m \text{ factors}} \otimes \Phi_C^{(l)}(\bar{\rho}^{(l)}). \tag{15.86}$$

Inserting this into the definition of $H(\Phi(\bar{\rho}))$ and using the operator monotonicity of the logarithm and the fact that (γ_i) is the equilibrium distribution, i. e. $\sum_{i\in\mathbb{I}} \gamma_i q_{ij} = q_j$, we obtain

$$H(\Phi_C^{(n)}(\rho^{(n)})) \geq mH(\Phi_C^{(n')}(\bar{\rho}^{(n')})) + H(\Phi_C^{(i)}(\bar{\rho}^{(l)})) + m\log\gamma. \tag{15.87}$$

On the other hand, by subadditivity,

$$H(\Phi_C^{(n)}(\rho_j^{(n)})) \leq \sum_{r=1}^{m} H(\Phi_C^{(n')}(\rho_{j_r}^{(n')})) + H(\Phi_C^{(l)}(\rho_{m+1}^{(l)})) \tag{15.88}$$

so that

$$\bar{\chi}_C^{(n)}(\{p_j^{(n)}, \Phi^{(n)}(\rho^{(n)})\}) \geq \frac{mn'}{n}\bar{\chi}_{n'} + \frac{m}{n}\log\gamma, \quad \forall C \in \mathcal{C}. \tag{15.89}$$

This proves the lemma. □

The existence of the following limit follows from Lemma 15.4.8.

Corollary 15.4.9. *Let $\Phi^{(\infty)} = \{\Phi^{(n)}\}_{n=1}^{+\infty}$ be the quantum channel with ergodic Markov memory described in* (15.5). *Then the following limit exists:*

$$\chi^*(\Phi) = \lim_{n\to+\infty} \frac{1}{n} \sup_{\{p_j^{(n)}, \rho_j^{(n)}\}} \chi(\{p_j^{(n)}, \Phi^{(n)}(\rho_j^{(n)})\}) \tag{15.90}$$

We follow the approach presented in Datta and Dorla [34] and [33] to prove Theorem 15.4.7 by establishing the following two propositions (Proposition 15.4.10 and Proposition 15.4.17), where its direct part is proved in Proposition 15.4.10 and Proposition 15.4.17 provides its weak converse part.

Proposition 15.4.10. *Let $\Phi^{(\infty)} = \{\Phi^{(n)}\}_{n=1}^{+\infty}$ be a quantum channel with general Markov memory defined by* (15.4)*, where $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$. For all $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that there exist at least $N = M_n \geq 2^{n(C^{\mathrm{mar}}(\Phi)-\epsilon)}$ product states $\tilde{\rho}_1^{(n)}, \tilde{\rho}_2^{(n)}, \ldots, \tilde{\rho}_N^{(n)}$ on $\mathbb{H}_A^{\otimes n}$ and positive operators $\mathbf{D}_1^{(n)}, \mathbf{D}_2^{(n)}, \ldots, \mathbf{D}_N^{(n)}$ on $\mathbb{H}_B^{\otimes n}$ such that $\sum_{i=1}^N \mathbf{D}_i^{(n)} \leq \mathbf{1}$,*

$$\mathrm{tr}[\Phi^{(n)}(\tilde{\rho}_k^{(n)})\mathbf{D}_k^{(n)}] > 1 - \epsilon \tag{15.91}$$

for all $k = 1, \ldots, N$.

Remark 15.1. Note that the above proposition implies that a rate $R < C^{\mathrm{mar}}(\Phi)$ is achievable. This can be seen as follows: Given an $R < C^{\mathrm{mar}}(\Phi)$, choose $\epsilon > 0$ such that $R < C^{\mathrm{mar}}(\Phi) - \epsilon$. Then Theorem 15.4.7 guarantees the existence of codes $\mathfrak{C}^{(n)}$ of length n and size,

$$N_n \geq 2^{n(C^{\mathrm{mar}}(\Phi)-\epsilon)} \geq 2^{nR},$$

with codewords given by product states $\rho_j^{(n)}$, and POVM elements $\{\mathbf{D}_j^{(n)}\}$, for which the probability of error, $\bar{\mathbb{P}}_{\mathrm{err}}^{(n)}$, can be made arbitrarily small, for each $j \in \{1, 2, \ldots, N_n\}$ and n large enough. Hence, the rate R is achievable.

Proof of Proposition 15.4.10. Given $\delta > 0$, we can choose m_0 large enough such that

$$\operatorname{tr}[\tilde{\Pi}_C \Phi_C^{(m_0 M)}(\omega^{(m_0 M)})] > 1 - \delta \tag{15.92}$$

for all $C \in \mathcal{C}_{\text{aper}}$ and

$$\operatorname{tr}[\tilde{\Pi}_{C',i'} \Phi_{C',i'}^{(m_0 M)}(\omega^{(m_0 M)})] > 1 - \delta \tag{15.93}$$

for all $C' \in \mathcal{C}_{\text{per}}$ and $i' \in \mathcal{C}'$. Here, M is given by (15.65). The product state $\omega^{(m_0 M)}$, defined through (15.75), is used as a preamble to the input state encoding each message, and serves to distinguish between the different branches of the channel, i. e., between $\Phi_C, C \in \mathcal{C}_{\text{aper}}$ and $\Phi_{C',i}, C' \in \mathcal{C}_{\text{per}}$ and $i \in C'$.

If $\rho_k^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n})$ is a state encoding the kth classical message in the set Λ_n, then the kth codeword is given by the product state

$$\omega^{(m_0 M)} \otimes \rho_k^{(n)}.$$

First, we fix l_0 large enough, and an ensemble $\{p_j^{(l_0)}, \rho_j^{(l_0)}\}$ such that

$$\left| C^{\text{mar}}(\Phi) - \min_{C \in \mathcal{C}} \bar{\chi}_C(\{p_j^{(l_0)}, \rho_j^{(l_0)}\}) \right| < \frac{\epsilon}{6}. \tag{15.94}$$

Let $N = \tilde{N}(n)$ be the maximal number of product states $\tilde{\rho}_1^{(n)}, \ldots, \tilde{\rho}_N^{(n)}$ on $\mathbb{H}_A^{\otimes n}$ for which there exist positive operators $\mathbf{D}_1^{(n)}, \ldots, \mathbf{D}_N^{(n)}$ on $\mathbb{H}_B^{\otimes m_0 M} \otimes \mathbb{H}_B^{\otimes n}$ such that:

1. $\mathbf{D}_k^{(n)} = \sum_{C \in \mathcal{C}_{\text{aper}}} \tilde{\Pi}_i \otimes \mathbf{D}_{k,C}^{(n)} + \sum_{C' \in \mathcal{C}'_{\text{per}}} \sum_{i' \in C'} \tilde{\Pi}_{C',i'} \otimes \mathbf{D}_{k,i'}^{(n)}$ and $\sum_{k=1}^N \mathbf{D}_{k,C}^{(n)} \le \bar{\mathbf{P}}_{C,n}^{(n)}$; $\sum_{k=1}^N \mathbf{D}_{k,i'}^{(n)} \le \bar{\mathbf{P}}_{i',n}$, and
2. $\sum_{C \in \mathcal{C}_{\text{aper}}} \gamma_C \operatorname{tr}[(\tilde{\Pi}_C \otimes \mathbf{D}_{k,C}^{(n)}) \Phi_C^{(m_0 M + n)}(\omega^{(m_0 M)} \otimes \tilde{\rho}_k^{(n)})] + \sum_{C' \in \mathcal{C}_{\text{per}}} \sum_{i' \in C'} \gamma_{i'} \operatorname{tr}[(\tilde{\Pi}_{C',i'} \otimes \mathbf{D}_{k,C}^{(n)}) \Phi_{C',i'}^{(m_0 M + n)}(\omega^{(m_0 M)} \otimes \tilde{\rho}_k^{(n)})] > 1 - \epsilon$, and
3. $\sum_{C \in \mathcal{C}_{\text{aper}}} \gamma_C \operatorname{tr}[(\tilde{\Pi}_C \otimes \mathbf{D}_{k,C}^{(n)}) \Phi_C^{(m_0 M + n)}(\omega^{(m_0 M)} \otimes \tilde{\rho}_k^{(n)})] + \sum_{C' \in \mathcal{C}_{\text{per}}} \sum_{i' \in C'} \gamma_{i'} \operatorname{tr}[(\tilde{\Pi}_{C',i'} \otimes \mathbf{D}_{k,C}^{(n)}) \Phi_{C',i'}^{(m_0 M + n)}(\omega^{(m_0 M)} \otimes \tilde{\rho}_k^{(n)})] \le 2^{-n(C^{\text{mar}}(\Phi) - \frac{1}{2}\epsilon)}$.

Note that, as in the ergodic case, we can append $\mathbf{1}^{(n-m_0)}$ to all POVM elements, to reduce the proof to the case $n = ml_0$. In the following, we therefore assume $n = ml_0$ for simplicity. The typical projection $\bar{P}_{C,n}$ for an aperiodic class is defined as before by Lemma 15.3.6. For a periodic class C', we define the typical spaces by interlacing those for the product channels $\Phi_i^{\otimes n}$ $(i \in C')$, as stated in the following lemma.

Lemma 15.4.11. *Let C' be a periodic class with period L. Given $\epsilon, \delta > 0$, there exists $m' \in \mathbb{N}$ such that for $m \ge m'$ there are subspaces $\bar{\mathbb{K}}^{(n)} \subset (\mathbb{H}_B^{(l_0)})^{\otimes} m$ $(i \in C'$, $n = ml_0)$, with projections $\bar{\mathbf{P}}_{i,n}$ such that*

$$\bar{\mathbf{P}}_{i,n} \Phi_{C',i}(\rho_{l_0}^{\otimes m}) \bar{\mathbf{P}}_{i,n} \le 2^{-m[H_{C'} - \frac{\epsilon}{4}]}, \tag{15.95}$$

where

$$H_{C'} = \frac{1}{L}\sum_{i=0}^{L-1} H(\Phi_{C',i}^{(l_0)}(\bar{\rho}^{(l_0)}))$$

and

$$\mathrm{tr}[\Phi_{C',i}^{(n)}((\bar{\rho}^{(l_0)})^{\otimes} m)\bar{\mathbf{P}}_{i,n}] > 1-\delta^2. \tag{15.96}$$

Proof. We simply let $\bar{K}_{i,\epsilon}^{(n)}$ be the subspace spanned by the vectors $|\psi_{i,k_1}\rangle \otimes \cdots \otimes |\psi_{i+l_0(n-1),k_n}\rangle$, where $|\psi_{i,k}\rangle$ is an eigenvector of $\Phi_{C',i}^{(l_0)}$ and $|\psi_{i,k_1}\rangle \otimes |\psi_{i,k_{L+1}}\rangle \otimes \cdots \otimes |\psi_{i,k_{|(n-1)/L|L+1}}\rangle$ belongs to the typical space for $\Phi_{C,i}^{(l_0)}(\bar{\rho}^{(l_0)})$, $|\psi_{i+1,k_2}\rangle \otimes |\psi_{i+1,k_{L+1}}\rangle \otimes \cdots \otimes |\psi_{i+1,k_{|(n-1)/L|L+2}}\rangle$ to that of $\Phi_{C,i+1}^{(l_0)}(\bar{\rho}^{(l_0)})$, etc. This proves the lemma. □

Similarly, we have the following lemma.

Lemma 15.4.12. *Let C' be a periodic class with period L. Given $i \in C'$, and a sequence $\vec{j} \in \{1,\ldots,j\}^{\times m}$, let $\mathbf{P}_{i,j}^{(n)} = \mathbf{P}_{(C',i),j}^{(n)}$ by the projection onto the subspace of $\mathbb{H}_B^{\otimes n}$ spanned by the vectors*

$$\Phi_{C',i}^{(n)}(\rho_{\vec{j}}^{(l_0)}) = \bigotimes_{r=1}^{m} \Phi_{C',i+(r-1)l_0}^{(l_0)}(\rho_{j_r}^{(l_0)}),$$

with eigenvalues $\lambda_{\vec{j},\vec{k}} = \prod_{r=1}^{m} \lambda_{i+(r-1)l_0,i_r,k_r}$ such that

$$\left|\frac{1}{m}\log\lambda_{\vec{j},\vec{k}} + \bar{H}_{C'}\right| < \frac{\epsilon}{4}, \tag{15.97}$$

where

$$\bar{H}_{C'} = \lim_{m\to+\infty}\frac{1}{mL}\sum_{i'\in C'}\sum_{\vec{j}} p_{\vec{j}}^{(n)} H\big((\Phi_{i'}\otimes\cdots\otimes\Phi_{i'+ml_0+1})(\rho_{j_1}^{(l_0)}\otimes\cdots\otimes\rho_{j_m}^{(l_0)})\big)$$

For any $\delta > 0$, there exists m_2' such that for $m \ge m_2'$,

$$\mathbb{E}\left(\mathrm{tr}\left[\Phi_{C',i}^{(ml_0)}\left(\bigotimes_{r=1}^{m}\rho_{j_r}^{(l_0)}\right)\mathbf{P}_{i,\vec{j}}^{(n)}\right]\right) > 1-\delta^2.$$

The proof of Proposition 15.4.10 is continued as follows.

For each $c = C$ or $c = (C', i')$ with $i' \in C' \in \mathcal{C}_{\mathrm{per}}$, and $\vec{j} = (j_1,\ldots,j_m)$, we define

$$\mathbf{V}_{c,\vec{j}}^{(n)} = \left(\bar{\mathbf{P}}_c^{(n)} - \sum_{k=1}^{N}\mathbf{D}_{k,c}^{(n)}\right)^{1/2} \bar{\mathbf{P}}_c^{(n)}\mathbf{P}_{c,\vec{j}}^{(n)}\bar{\mathbf{P}}_c^{(n)}\left(\bar{\mathbf{P}}_c^{(n)} - \sum_{k=1}^{N}\mathbf{D}_{k,c}^{(n)}\right)^{1/2} \tag{15.98}$$

Clearly, $\mathbf{V}^{(n)}_{c,\vec{j}} \leq \bar{\mathbf{P}}^{(n)}_{c} - \sum_{k=1}^{N} \mathbf{D}^{(n)}_{k,c}$. Put

$$\mathbf{V}^{(n)}_{\vec{j}} = \sum_{C\in\mathcal{C}_{\text{aper}}} \tilde{\Pi}_i \otimes \mathbf{V}^{(n)}_{C,\vec{j}} + \sum_{C'\in\mathcal{C}_{\text{per}}} \sum_{i'\in C'} \tilde{\Pi}_{C',i'} \otimes \mathbf{V}^{(n)}_{(C',i'),\vec{j}}. \tag{15.99}$$

This is a candidate for an additional measurement operator, $\mathbf{D}^{(n)}_{N+1}$, for Bob with corresponding input state $\tilde{\rho}^{(n)}_{N+1} = \rho^{(n)}_{\vec{j}} = \rho_{j_1} \otimes \rho_{j_2} \otimes \cdots \rho_{j_n}$. Clearly, the condition (1) given under (15.94) is satisfied and we also have the following lemma.

Lemma 15.4.13. *The following inequality holds:*

$$\begin{aligned}&\sum_{C\in\mathcal{C}_{\text{aper}}} \gamma_C \operatorname{tr}\big[(\tilde{\Pi}_C \otimes \mathbf{V}^{(n)}_{C,\vec{j}})\Phi_C^{\otimes m'M+n}(\omega^{(m'M)} \otimes (\bar{\rho}^{(l_0)})^{\otimes[n/l_0]})\big]\\ &\quad + \sum_{C'\in\mathcal{C}_{\text{per}}} \sum_{i'\in C'} \gamma_{i'} \operatorname{tr}\big[(\tilde{\Pi}_{C',i'} \otimes \mathbf{V}_{(C',i'),\vec{j}})\Phi_C^{\otimes m'M+n}(\omega^{(m'M)} \otimes (\bar{\rho}^{(l_0)})^{\otimes[n/l_0]})\big]\\ &\leq 2^{-n(C^{\text{mar}}(\Phi)-\frac{2}{3}\epsilon)},\end{aligned} \tag{15.100}$$

with $\gamma_{i'} = 1/L(C')$ *for* $i' \in C' \in \mathcal{C}_{\text{per}}$.

Proof. Writing $\bar{\sigma}^{(n)}_C = \Phi^{(n)}_C(\bar{\rho}^{(n)})$, by the proof of Lemma 15.3.8, the following inequality holds for an aperiodic class C, for n large enough:

$$\operatorname{tr}\big[\bar{\sigma}^{(n)}_C \mathbf{V}^{(n)}_{C\vec{j}}\big] \leq 2^{-n[\bar{\chi}_C - \frac{\epsilon}{2}]}, \tag{15.101}$$

where $\bar{\chi}_C = \bar{\chi}^{(l_0)}_C$ is given by (15.76), for the maximizing ensemble, with $n = l_0$. Then

$$\begin{aligned}&\sum_{C\in\mathcal{C}_{\text{aper}}} \gamma_C \operatorname{tr}\big[(\tilde{\Pi}_C \otimes \mathbf{V}^{(n)}_{C,\vec{j}})\Phi_C^{(m_0M+n)}(\omega^{(m_0M)} \otimes \bar{\rho}^{(n)})\big]\\ &\quad\leq \sum_{C\in\mathcal{C}_{\text{aper}}} \gamma_C \operatorname{tr}\big[\bar{\sigma}^{(n)}_C \mathbf{V}^{(n)}_{C,\vec{j}}\big]\\ &\quad\leq \sum_{C\in\mathcal{C}_{\text{aper}}} \gamma_C 2^{-n[\bar{\chi}_C - \frac{1}{2}\epsilon]}\end{aligned} \tag{15.102}$$

where we used the fact that $\tilde{\Pi}_C \leq \mathbf{1}$ and (15.100).

Similarly, for $i' \in C' \in \mathcal{C}_{\text{per}}$, denoting $\mathbf{Q}_{n,i'} = \sum_{k=1}^{N} \mathbf{D}^{(n)}_{k,i'}$, we have using Lemma 15.4.11,

$$\bar{\mathbf{P}}_{i',n}\Phi^{(n)}_{C',i'}(\rho_{l_0}^{\otimes m})\bar{\mathbf{P}}_{i',n} \leq 2^{-m[H_{C'} - \frac{\epsilon}{4}]},$$

and hence,

$$\begin{aligned}\operatorname{tr}[\bar{\sigma}^{(n)}_{C',i'}\mathbf{V}^{(n)}_{i',\vec{j}}] &= \operatorname{tr}[\bar{\sigma}^{(n)}_{C',i'}(\bar{\mathbf{P}}^{(n)}_{C',i'}-\mathbf{Q}_{n,i'})^{1/2}\bar{P}^{(n)}_{C',i'}\mathbf{P}^{(n)}_{i',\vec{j}}\bar{P}^{(n)}_{C',i'}(\bar{\mathbf{P}}^{(n)}_{C',i'}-\mathbf{Q}_{n,i'})^{1/2}]\\ &\le 2^{-m[H_{C'}-\frac{\epsilon}{4}]}\operatorname{tr}[(\bar{\mathbf{P}}^{(n)}_{C',i'}-\mathbf{Q}_{n,i'})^{1/2}\mathbf{P}^{(n)}_{i',\vec{j}}(\bar{\mathbf{P}}^{(n)}_{C',i'}-\mathbf{Q}_{n,i'})^{1/2}]\\ &\le 2^{-m[H_{C'}-\frac{\epsilon}{4}]}\operatorname{tr}[\mathbf{P}^{(n)}_{i',\vec{j}}] \le 2^{-n[\frac{1}{l_0}(H_{C'}-\bar{H}_{C'})-\frac{\epsilon}{4}]} \le 2^{-n[\bar{\chi}^{(l_0)}_{C'}-\frac{1}{2}]},\end{aligned} \tag{15.103}$$

where

$$\bar{\chi}^{(l_0)}_{C'} = \frac{1}{l_0L}\sum_{i\in C'}(H(\Phi^{(l_0)}_{C',i}*\bar{\rho}^{(l_0)})-\bar{H}_i).$$

In the above, we have used the fact that $\operatorname{tr}[\mathbf{P}^{(n)}_{i',\vec{j}}] \le 2^{m\bar{H}_{C'}+\frac{\epsilon}{4}}$, which is a standard consequence of Lemma 15.4.12. We obtain the last line of (15.103) by using the subadditivity of the von Neumann entropy. Summing (15.103) over i' and C', and adding to the bound for $C\in\mathcal{C}_{\text{aper}}$, yields the following bound:

$$\begin{aligned}&\sum_{C\in\mathcal{C}_{\text{aper}}}\gamma_C\operatorname{tr}[(\tilde{\Pi}_C\otimes\mathbf{V}^{(n)}_{C,\vec{j}})\Phi_C^{\otimes m'M+n}(\omega^{(m'M)}\otimes(\bar{\rho}^{(l_0)})^{\otimes[n/l_0]})]\\ &\quad+\sum_{C'\in\mathcal{C}_{\text{per}}}\sum_{i'\in C'}\gamma_{i'}\operatorname{tr}[(\tilde{\Pi}_{C',i'}\otimes\mathbf{V}_{(C',i'),\vec{j}})\Phi_C^{\otimes m'M+n}(\omega^{(m'M)}\otimes(\bar{\rho}^{(l_0)})^{\otimes[n/l_0]})]\\ &\le\sum_{C\in\mathcal{C}_{\text{aper}}}\gamma_C2^{-n[\bar{\chi}^{(l_0)}_C-\frac{\epsilon}{2}]}+\sum_{C'\in\mathcal{C}_{\text{per}}}\sum_{i\in C'}\gamma_i2^{-n[\bar{\chi}^{(l_0)}_{C'}-\frac{\epsilon}{2}]}\end{aligned}$$

Now by (15.94),

$$C^{\text{mar}}(\Phi)\le\min_{C\in\mathcal{C}}\bar{\chi}^{(l_0)}_C+\frac{\epsilon}{6},$$

and hence,

$$2^{-n[\bar{\chi}^{(l_0)}_{C'}-\frac{\epsilon}{2}]}\le 2^{-n[C^{\max}(\Phi)-\frac{2\epsilon}{3}]},\quad\forall C\in\mathcal{C},$$

and, therefore, (15.100). This proves the lemma. □

By maximality of N, it now follows that the condition (2) above cannot hold, that is, we have the following.

Corollary 15.4.14.

$$\begin{aligned}&\sum_{C\in\mathcal{C}_{\text{aper}}}\gamma_C\mathbb{E}(\operatorname{tr}[(\tilde{\Pi}_C\otimes\mathbf{V}^{(n)}_{C,\vec{j}})\Phi_C^{\otimes m'M+n}(\omega^{(m'M)}\otimes(\bar{\rho}^{(l_0)})^{\otimes[n/l_0]})])\\ &\quad+\sum_{C'\in\mathcal{C}_{\text{per}}}\sum_{i'\in C'}\gamma_{i'}\mathbb{E}(\operatorname{tr}[(\tilde{\Pi}_{C',i'}\otimes\mathbf{V}_{(C',i'),\vec{j}})\Phi_C^{\otimes m'M+n}(\omega^{(m'M)}\otimes(\bar{\rho}^{(l_0)})^{\otimes[n/l_0]})])\\ &\le 1-2\epsilon.\end{aligned} \tag{15.104}$$

We also need the following lemma.

Lemma 15.4.15. *Assume $\eta' \geq 3\delta$. Then for n large enough,*

$$\begin{aligned}\sum_{C\in\mathcal{C}_{\text{aper}}} \gamma_C \operatorname{tr}[(\tilde{\Pi}_C \otimes \mathbf{V}^{(n)}_{C,\vec{j}})\Phi_C^{\otimes m'M+n}(\omega^{(m'M)} \otimes (\bar{\rho}^{(l_0)})^{\otimes[n/l_0]})] \\ + \sum_{C'\in\mathcal{C}_{\text{per}}} \sum_{i'\in C'} \gamma_{i'} \operatorname{tr}[(\tilde{\Pi}_{C',i'} \otimes \mathbf{V}_{(C',i'),\vec{j}})\Phi_C^{\otimes m'M+n}(\omega^{(m'M)} \otimes (\bar{\rho}^{(l_0)})^{\otimes[n/l_0]})] \\ > 1 - \eta'. \qquad (15.105)\end{aligned}$$

Proof. This is a simple consequence of Lemma 15.3.9 and its analogue of periodic classes, together with (15.93) and (15.94). This proves the lemma. □

The next lemma is analogous to Lemma 15.4.11 and is omitted here.

Lemma 15.4.16. *Assume $\eta' < \frac{1}{3}\epsilon$ and write*

$$\mathbf{Q}^{(n)}_{n,C} = \sum_{k=1}^{N} \mathbf{D}^{(n)}_{k,C} \quad (C \in \mathcal{C}_{\text{aper}}) \quad \textit{and} \quad \mathbf{Q}^{(n)}_{n,i'} = \sum_{k=1}^{N} \mathbf{D}^{(n)}_{k,i'} \quad (i' \in C' \in \mathcal{D}_{\text{per}}). \qquad (15.106)$$

Then for n large enough,

$$\begin{aligned}\sum_{C\in\mathcal{C}_{\text{aper}}} \gamma_C \operatorname{tr}[(\tilde{\Pi}_C \otimes \mathbf{Q}_{n,C})\Phi_C^{\otimes(m'M+n)}(\omega^{(m'M)} \otimes \rho^{(n)}_{\vec{j}})] \\ + \sum_{C'\in\mathcal{C}_{\text{per}}} \sum_{i'\in C'} \gamma_{i'} \operatorname{tr}[(\tilde{\Pi}_{C',i'} \otimes \mathbf{Q}_{n,i'})\Phi_C^{\otimes(m'M+n)}(\omega^{(m'M)} \otimes \rho^{(n)}_{\vec{j}})] \\ > (\eta')^2 \qquad (15.107)\end{aligned}$$

Back to the proof of Proposition 15.4.10, it now follows, as before, that for n large enough, $\tilde{N}(n) \geq (\eta')^2 2^{n[C^{\max}(\Phi)-\frac{2\epsilon}{3}]}$. We take the following states as codewords:

$$\rho_k^{(m_0M+n)} = \omega^{(m_0+M)} \otimes \tilde{\rho}_k^{(n)}$$

For n sufficiently large, we then have

$$N = N_{n+m_0M} = \tilde{N}(n) \geq (\eta')^2 2^{n[C^{\max}(\Phi)-\frac{2}{3}\epsilon]} \geq 2^{(m_0M+n)[C^{\max}(\Phi)-\epsilon]}.$$

To complete the proof of Proposition 15.4.10, we note as before for large enough n that $\tilde{N}(n) \geq (\eta')^2 2^{n[C^{\text{mar}}(\Phi)-\frac{3}{4}\epsilon]}$. We take the following states as codewords:

$$\rho_k^{(mL+n)} = \omega^{(mL)} \otimes \tilde{\rho}_k^{(n)}, \qquad (15.108)$$

where $\omega^{(mL)}$ is the preamble defined by (16.13). Consequently, for sufficiently large n, we have

$$N_{mL+n} = \tilde{N}(n) \geq (\eta')^2 2^{n[C^{lm}_{\text{prod}}(\Phi)-\frac{3}{4}\epsilon]} \geq 2^{(mL+n)(C^{lm}_{\text{prod}}(\Phi)-\epsilon)}.$$

To complete the proof, we need to show that the set $\{\mathbf{D}_k^{(n)}\}$ satisfies (15.96). But this follows immediately from condition (2) stated earlier:

$$\begin{aligned}
&\operatorname{tr}[\Phi^{(m_0M+n)}(\rho_k^{(m_0M+n)})\mathbf{D}_k^{(n)}] \\
&\quad = \sum_{C\in\mathcal{C}_{\text{aper}}} \gamma_C \operatorname{tr}[\Phi_C^{(m_0M+n)}(\omega^{(m_0M)} \otimes \tilde{\rho}_k^{(n)})\mathbf{D}_k^{(n)}] \\
&\qquad + \sum_{C'\in\mathcal{C}_{\text{per}}} \sum_{i\in C'} \gamma_i \operatorname{tr}[\Phi_{C',i}^{(m_0M+n)}(\omega^{(m_0M)} \otimes \tilde{\rho}_k^{(n)})\mathbf{D}_k^{(n)}] \\
&\quad = \sum i \operatorname{tr}[(\tilde{\Pi}_C \otimes \mathbf{D}_{k,C}^{(n)})\Phi^{\otimes mL}(\omega^{(mL)})] \operatorname{tr}[\Phi_{C',i}^{\otimes n}(\omega^{(m_0M)} \otimes \tilde{\rho}_k^{(n)})] \\
&\qquad + \sum_{C'\in\mathcal{C}_{\text{per}}} \sum_{i\in C'} \gamma_i \operatorname{tr}[(\tilde{\Pi}_i \otimes \mathbf{D}_{k,i}^{(n)})\Phi_{C',i}^{(m_0M+n)}(\omega^{(m_0M)} \otimes \tilde{\rho}_k^{(n)})] \\
&\quad \geq 1-\epsilon.
\end{aligned}$$

This proves Proposition 15.4.10. □

The following proposition proves that it is impossible for Alice to transmit classical messages reliably to Bob through the channel Φ with general Markov memory defined by (15.41) at a rate $R > C^{\text{mar}}(\Phi)$.

Proposition 15.4.17 (Weak converse of Theorem 15.4.7). *In the sense that the probability of error does not tend to zero asymptotically as the length of the code increases, for any code with rate $R > C^{\text{mar}}(\Phi)$.*

Proof. To prove the weak converse, suppose that Alice encodes messages labeled by $\lambda \in \Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{M_n}\}$ by the state $\rho_\lambda^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n})$. Let the corresponding outputs for the class $C \in \mathcal{C}$ of the channel be denoted by $\sigma_{\lambda,C}^{(n)}$, i. e.,

$$\sigma_{\lambda,C}^{(n)} = \Phi_C^{(n)}(\rho_\lambda^{(n)}). \tag{15.109}$$

Further define

$$\bar{\sigma}_C^{(n)} = \frac{1}{|\Lambda_n|} \sum_{\lambda\in\Lambda_n} \sigma_{\lambda,C}^{(n)}.$$

Let Bob's POVM elements corresponding to the codewords $\rho_\lambda^{(n)}$ be denoted by $\mathbf{D}_\lambda^{(n)}$, $\lambda \in \Lambda_n$. We may assume that Alice's messages are produced uniformly at random from the set Λ_n. Then Bob's average probability of error is given by

$$\bar{\mathbb{P}}_{\text{err},C}^{(n)} := 1 - \frac{1}{|\Lambda_n|} \sum_{\lambda\in\Lambda_n} \operatorname{tr}[\Phi_C^{(n)}(\rho_\lambda^{(n)})\mathbf{D}_\lambda^{(n)}]. \tag{15.110}$$

So that

$$\bar{\mathbb{P}}_{\text{err}}^{(n)} = \sum_{C\in\mathcal{C}} \gamma_C \bar{\mathbb{P}}_{\text{err},C}^{(n)}. \tag{15.111}$$

Let $X^{(n)}$ be a random variable with a uniform distribution over the set Λ_n, characterizing the classical message sent by Alice to Bob. Let $Y_C^{(n)}$ be the random variable corresponding to Bob's inference of Alice's message, when the codeword $\rho_i^{(n)}$ is transmitted through class C. It is defined by the conditional probabilities

$$\Pr[Y_C^{(n)} = y | X^{(n)} = \lambda] = \operatorname{tr}[\Phi_C^{\otimes n}(\rho_\lambda^{(n)})\mathbf{D}_y^{(n)}]. \tag{15.112}$$

By the following Fano's inequality (see also (16.36)),

$$\begin{aligned} &h(\bar{\mathbb{P}}_{\text{err},C}^{(n)}) + \bar{\mathbb{P}}_{\text{err},C}^{(n)} \log(|\Lambda_n| - 1) \\ &\quad \geq H(X^{(n)} | Y_C^{(n)}) \\ &\quad = H(X^{(n)}) - H(X^{(n)} \| Y_C^{(n)}) \end{aligned} \tag{15.113}$$

Here, $h_2(\cdot)$ denotes the binary entropy and $H(\cdot)$ denotes the Shannon entropy. Using the Holevo bound and the subadditivity of the von Neumann entropy (see Proposition 7.2.8), we have

$$\begin{aligned} &H(X^{(n)} \| Y_C^{(n)}) \\ &\quad \leq H\bigg(\frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} \Phi_C^{\otimes n}(\rho_\lambda^{(n)})\bigg) - \frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} H(\Phi_C^{\otimes n}(\rho_\lambda^{(n)})) \\ &\quad = n\bar{\chi}_C\bigg(\bigg\{\frac{1}{|\Lambda_n|}, \rho_\lambda^{(n)}\bigg\}_{\lambda\in\Lambda_n}\bigg), \end{aligned} \tag{15.114}$$

where $\bar{\chi}_C(\{\frac{1}{|\Lambda_n|}, \rho_\lambda^{(n)}\}_{\lambda\in\Lambda_n})$ is given in (15.78).

For $C \in \mathcal{C}_{\text{per}}$ with period L,

$$\begin{aligned} &H(X^{(n)} \| Y_C^{(n)}) \\ &\quad \leq H\bigg(\frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} \frac{1}{L}\sum_{i\in C} \Phi_{C,i}^{(n)}(\rho_\lambda^{(n)})\bigg) - \frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} H\bigg(\frac{1}{L}\sum_{i\in C} \Phi_{C,i}^{(n)}(\rho_\lambda^{(n)})\bigg) \\ &\quad = \frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} H\bigg(\frac{1}{L}\sum_{i\in C} \Phi_{C,i}^{(n)}(\rho_\lambda^{(n)}) \| \frac{1}{|\Lambda_n|}\sum_{y\in\Lambda_n} \frac{1}{L}\sum_{i\in C} \Phi_{C,i}^{(n)}(\rho_y^{(n)})\bigg) \\ &\quad \leq \frac{1}{|\Lambda_n| L}\sum_{\lambda\in\Lambda_n}\sum_{i\in C} H(\Phi_{C,i}^{(n)}\bigg(\rho_\lambda^{(n)} \| \frac{1}{|\Lambda_n|}\sum_{y\in\Lambda_n} \Phi_{C,i}^{(n)}(\rho_y^{(n)})\bigg) \\ &\quad = \frac{1}{L}\sum_{i\in C} \chi_{C,i}^{(n)}\bigg(\bigg\{\frac{1}{|\Lambda_n|}, \rho_\lambda^{(n)}\bigg\}_{\lambda\in\Lambda_n}\bigg) \\ &\quad = n\bar{\chi}_C^{(n)}\bigg(\bigg\{\frac{1}{|\Lambda_n|}, \rho_\lambda^{(n)}\bigg\}_{\lambda\in\Lambda_n}\bigg). \end{aligned} \tag{15.115}$$

In the above, we use the convexity of the relative entropy $H(\sigma\|\omega) := \mathrm{tr}[\sigma(\log\sigma - \log\omega)$, for quantum σ and ω.

Therefore, for any class C, we have the upper bound

$$H(X^{(n)}\|Y_C^{(n)}) \le n\bar{\chi}_C^{(n)}\left(\left\{\frac{1}{|\Lambda_n|}, \rho_\lambda^{(n)}\right\}_{\lambda\in\Lambda_n}\right).$$

Inserting this into Fano's inequality, (15.113), now yield

$$h(\bar{\mathbb{P}}_{\mathrm{err},C}^{(n)}) + \bar{\mathbb{P}}_{\mathrm{err},C}^{(n)} \log(|\Lambda_n|) \ge \log(|\Lambda_n|) - n\bar{\chi}_C^{(n)}\left(\left\{\frac{1}{|\Lambda_n|}, \rho_\lambda^{(n)}\right\}_{\lambda\in\Lambda_n}\right).$$

However, since

$$C^{\mathrm{mar}}(\Phi) \ge \min_{C\in\mathcal{C}} \bar{\chi}_C^{(n)}\left(\left\{\frac{1}{|\Lambda_n|}, \rho_\lambda^{(n)}\right\}_{\lambda\in\Lambda_n}\right), \tag{15.116}$$

and $R = \frac{1}{n}\log|\Lambda_n| > C^{\mathrm{mar}}(\Phi)$, there must be at least one class C such that

$$\bar{\mathbb{P}}_{\mathrm{err},C}^{(n)} \ge 1 - \frac{C^{\mathrm{mar}}(\Phi) + 1/n}{R} > 0. \tag{15.117}$$

We conclude from (15.111) and (15.117)) that

$$\bar{\mathbb{P}}_{\mathrm{err}}^{(n)} \ge \left(1 - \frac{C^{\mathrm{mar}}(\Phi) + 1/n}{R}\right) \min_{C\in\mathcal{C}} \gamma_C. \tag{15.118}$$

This proves the theorem. □

16 Channels with long-term memory

Datta and Dolas [31] (see also Dorlas and Morgan [42] and Datta–Suhov–Dorlas [35]) investigated the product state classical capacity as well as entanglement assisted classical capacity of a convex combination of memoryless channels that are the subjects of exposition in this chapter.

16.1 The model

The model obtained in [31, 42] and [35] for quantum channel with long term memory is described below.

Given a collection of M memoryless channels $\{\Phi_i\}_{i=1}^M$ with common input Hilbert space $\mathbb{H}_A$ and the common output Hilbert space $\mathbb{H}_B$, a convex combination (or average) of these channels is defined by the map

$$\Phi(\rho_A) = \sum_{i=1}^{M} p_i \Phi_i(\rho_A), \quad \forall \rho_A \in \mathcal{S}(\mathbb{H}_A), \tag{16.1}$$

where $\{p_i\}_{i=1}^M$ is probability distribution over choices of memoryless channels $\{\Phi_i\}_{i=1}^M$.

The n-use of Φ defined in (16.1) yields a channel $\Phi^{(n)} : \mathcal{S}(\mathbb{H}_A^{\otimes n}) \to \mathcal{S}(\mathbb{H}_B^{\otimes n})$ defined by

$$\Phi^{(n)}(\rho_A^{(n)}) = \sum_{i=1}^{M} p_i \Phi_i^{\otimes n}(\rho_A^{(n)}), \quad \forall \rho_A^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n}) \tag{16.2}$$

for $n = 1, 2, \ldots$. Note that the probability p_i in the above expression remains unchanged even after n uses of the channel. This is because this model implicitly assumes that once the channel Φ_i is chosen then it will be used for the n-consecutive times. Therefore, the "long term memory" is named. This implicit assumption will be mathematical expressed later.

The convex combination channel Φ defined in (16.1) will have a long-term memory in the sense that it is the simplest of nonforgetful channels. The channel with long-term memory is a channel with Markovian memory, which is aperiodic but not irreducible, using Markov chain terminologies in Section 15.1. This can be seen as follows. An n-use of a quantum channel with Markovian correlated noise given by a CPTP map $\Phi^{(n)} : \mathfrak{B}(\mathbb{H}_A^{\otimes n}) \to \mathfrak{B}(\mathbb{H}_B^{\otimes n})$ (see (15.5) for the definition of channels with Markov memory) can be stated as

$$\Phi^{(n)}(\rho_A^{(n)}) = \sum_{i_1, i_2, \ldots, i_n = 1}^{M} \gamma_{i_1} q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{n-1} i_n} (\Phi_{i_1} \otimes \cdots \otimes \Phi_{i_n})(\rho_A^{(n)}), \tag{16.3}$$

https://doi.org/10.1515/9783110788105-016

where (i) $q_{ij} = \Pr[X_{n+1} = j|X_n = i]$ denotes the elements of the transition matrix of a discrete time Markov chain $\{X_n\}_{n=0}^{+\infty}$ with a finite state space $\mathbb{I} = \{1, 2, \dots, M\}$; (ii) $\gamma_{i_1} = \lim_{n\to+\infty} \Pr[X_n = i_1]$ is the stationary distribution of $\{X_n\}_{n=0}^{+\infty}$ at $i_1 \in \mathbb{I}$ and (iii) for each $i \in \mathbb{I}$, $\Phi_i : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ is a memoryless channel. Casting the channel defined by (16.2) in this form (16.3) yields $q_{ij} = \delta_{ij}$ and $\gamma_i = p_i$. Hence, the transition matrix of the Markov chain, in this case, is the identity matrix. In other words, once a particular branch $i_1 \in \mathbb{I} = \{1, 2, \dots, M\}$ has been chosen with the probability γ_{i_1}, the successive inputs are sent through this branch. Transition between different branches (which correspond to the different states of the Markov chain) is not permitted. In this case, the Markov chain $\{X_n\}_{n=0}^{+\infty}$ that governs the switching among the M branches of memoryless channels Φ_i, $i = 1, 2, \dots, M$ is therefore aperiodic but not irreducible. Hence, the channel Φ defined by (16.1) possess long-term memory and does not lie in the class of forgetful channels explored in Section 14.4.

This chapter details developments of product state capacity $C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi)$ and entanglement assisted classical capacity $C^{\mathrm{lm}}_{\mathrm{ea}}(\Phi)$ for both infinite-dimensional unconstrained and energy constrained long term memory channel Φ. These developments extend finite-dimensional results due originally to Datta and Dolas [31] and Datta–Suhov–Dorlas [35], respectively.

It is an important issue to consider infinite-dimensional and yet constrained quantum communication systems. When applying the protocol of classical and entanglement-assisted communication to infinite-dimensional quantum channels one has to impose certain constraints on the input states, in particular, the constrained on each branch of the memoryless channels Φ_i, $i = 1, 2, \dots, M$, that constitute the channel Φ with long-term memory defined by (16.1).

16.2 Classical product state capacity

We consider the scenario in which the sender, Alice, wants to send a sequence classical messages to the receiver, Bob, through repeated usage of the channel $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ described by (16.1) as follows. For each $n \in \mathbb{N}$, Alice will first choose a classical message $\lambda \in \Lambda_n = \{\lambda_1, \lambda_2, \dots, \lambda_{N_n}\}$ at random, and encode the classical message λ into the codeword $\rho^{(n)}_{\lambda,A} \in \mathcal{S}(\mathbb{H}^{(n)}_A)$, where we have assumed that $\mathbb{H}_{\lambda,A} = \mathbb{H}_A$ for all $\lambda \in \Lambda$. She then will send the codeword through n-use of the channel Φ that consists of M memoryless branches Φ_i. Since one of the memoryless channels Φ_i will be used with probability p_i for $i = 1, 2, \dots, M$. This introduces long-term memory and as a result the (product-state) capacity of the channel $\Phi^{(n)}$ is no longer given by the supremum of the Holevo quantity such as those given in Theorem 13.2.2. Instead, it was proved in [32] that the finite-dimensional product-state capacity $C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi)$ (where the superscript "lm" means "long-term memory" and the subscript "prod" means the input states ρ_j are product states instead of entangled ones) is given by

$$C_{\mathrm{prop}}^{\mathrm{lm}}(\Phi) = \sup_{\{p_i,\rho_i\}} \{\min\{\chi(\{p_i, \Phi_i(\rho_i)\}), i = 1, 2, \ldots, M\}\},$$

where $\chi(\cdot)$ is the Holevo χ function defined in (13.102).

The finite-dimensional product state capacity $C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)$ for long-term memory channel Φ is stated below.

Theorem 16.2.1 (Datta and Dorlas [32]). *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel with long term memory defined in* (16.1)*, where* $\dim(\mathbb{H}_A) < +\infty$ *and* $\dim(\mathbb{H}_B) < +\infty$*. Then the product state capacity $C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)$ of Φ is given by*

$$C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi) = \sup_{\{p_j,\rho_j\}} \{\min\{\chi(\{p_j, \Phi_i(\rho_j)\}), i = 1, 2, \ldots, M\}\}, \tag{16.4}$$

where (i) $\chi(\{p_j, \Phi_i(\rho_j)\})$ is the χ-function of $\{p_j, \Phi_i(\rho_j)\}$ and (ii) the supremum is taken over all finite ensembles of states $\rho_j \in \mathfrak{B}(\mathbb{H}_A)$ with probabilities p_j.

16.2.1 Infinite-dimensional classical capacity

The above theorem is extended to infinite-dimensional Hilbert spaces $\mathbb{H}_A$ and $\mathbb{H}_B$ and is stated as follows.

Theorem 16.2.2 (Infinite-dimensional case). *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a quantum channel with long term memory defined in* (16.1)*, where $\mathbb{H}_A$ and $\mathbb{H}_B$ are infinite-dimensional Hilbert spaces. Then the product state capacity $C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)$ of Φ is given by*

$$C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi) = \sup_{\{p_j,\rho_j\}} \{\min\{\chi(\{p_j, \Phi_i(\rho_j)\}), i = 1, 2, \ldots, M\}\}, \tag{16.5}$$

where (i) *$\chi(\{p_j, \Phi_i(\rho_j)\})$ is the χ-function of $\{p_j, \Phi_i(\rho_j)\}$ and* (ii) *the supremum is taken over all finite ensembles of states $\rho_j \in \mathfrak{B}(\mathbb{H}_A)$ with probabilities p_j.*

The above theorem will be proved in the following three parts, where construction of preambles, proofs of direct part and converse of Theorem 16.2.2 are provided in details.

A. Preambles with long-term memory

As noted earlier, a channel with long-term memory is the channel consisting of a convex combination of different branches of memoryless channels as originally investigated by Datta and Dorlas [32]. To distinguish between the different memoryless branches, Φ_i, $i = 1, 2, \ldots, M$, of the quantum channel Φ defined in (16.2), we proceed similar to those presented in Section 15.4.1 and add a preamble to the input state encoding each message in the set $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{N_n}\}$. This is given by an m-fold tensor product of a suitable state as described below. Let us first sketch the idea behind adding such a preamble which distinguishes different branches of the channel Φ. The basic idea is based on the work by Helstrom [65] who showed that two states σ_1 and σ_2, occurring with a priori probabilities γ_1 and γ_2, respectively, can be distinguished, with an asymptotically vanishing probability of error, if a suitable collective measurement is performed on the m-fold tensor products $\sigma_1^{\otimes m}$ and $\sigma_2^{\otimes m}$, for a sufficiently large $m \in \mathbb{N}$. The relevant projection operators, which we denote by Π_+ and Π_-, are the orthogonal projections onto the positive and negative eigenspaces of the difference operator $\mathbf{A}_m = \gamma_1\sigma_1^{\otimes m} - \gamma_2\sigma_2^{\otimes m}$.

In the construction of preamble (see Datta and Dorlas [31]), we generalize Helstrom's result to distinguish between the different branches Φ_i. If the preamble is given by a state ω, then by using Helstrom's result, we can construct a POVM which distinguishes between the output states $\sigma_i^{\otimes m} := (\Phi_i(\omega))^{\otimes m}$ corresponding to the different branches Φ_i, $i = 1, 2, \ldots, M$. The outcome of this POVM measurement would in turn serve to determine which branch of the channel is being used for transmission.

We first note that we may assume that all branches Φ_i are different. Indeed, otherwise we do not need to distinguish them and can introduce a compound probability for each set of identical branches. This assumption means that there exist states $\omega_{i,j}$ on $\mathbb{H}_A$ for each pair $1 \le i < j \le M$ such that $\Phi_i(\omega_{i,j}) \ne \Phi_j(\omega_{i,j})$. Introducing the fidelity of two states as in (6.1), we then have

$$F(\sigma, \sigma') = \operatorname{tr}\big[\sqrt{\sigma^{1/2}\sigma'\sigma^{1/2}}\big], \tag{16.6}$$

we obtain

$$\begin{aligned} &F(\Phi_i(\omega_{i,j}), \Phi_j(\omega_{i,j})) \\ &\quad = \operatorname{tr}\Big[\sqrt{(\Phi_i(\omega_{i,j}))^{1/2}\Phi_j(\omega_{i,j})(\Phi_i(\omega_{i,j}))^{1/2}}\Big] \\ &\quad \le f < 1 \quad (\text{by Proposition 6.1.1 because } \Phi_i(\omega_{i,j}) \ne \Phi_j(\omega_{i,j})) \end{aligned} \tag{16.7}$$

for all pairs (i, j). We now introduce, for any $m \in \mathbb{N}$ and $1 \le i < j \le M$, the difference operators

$$\mathbf{A}_{i,j}^{(m)} = p_i(\Phi_i(\omega_{i,j}))^{\otimes m} - p_j(\Phi_j(\omega_{i,j}))^{\otimes m}. \tag{16.8}$$

Let $\Pi_{i,j}^{\pm}$ be the orthogonal projections onto the eigenspaces of $\mathbf{A}_{i,j}^{(m)}$ corresponding to all nonnegative, and all negative eigenvalues, respectively.

Following the approach used in Datta and Dorlas [32], we prove the following series of lemmas.

Lemma 16.2.3. *Suppose that for a given* $\delta > 0$,

$$\left|\mathrm{tr}[|\mathbf{A}_{i,j}^{(m)}|] - (p_i + p_j)\right| \le \delta. \tag{16.9}$$

Then

$$\left|\mathrm{tr}[\Pi_{i,j}^{+}(\Phi_i(\omega_{i,j}))^{\otimes m}] - 1\right| \le \frac{\delta}{2p_i} \tag{16.10}$$

and

$$\left|\mathrm{tr}[\Pi_{i,j}^{-}(\Phi_j(\omega_{i,j}))^{\otimes m}] - 1\right| \le \frac{\delta}{2p_i}. \tag{16.11}$$

Proof. Write $\mathbf{A} = \mathbf{A}_{i,j}^{(m)}$ and $\Pi^{\pm} = \Pi_{i,j}^{\pm}$ for notational simplicity. We first note that

$$\begin{aligned}\mathrm{tr}[\Pi^{\pm}\mathbf{A}] &= \frac{1}{2}\,\mathrm{tr}[\mathbf{A} \pm (\Pi^{+} - \Pi^{-})\mathbf{A}] \\ &= \frac{1}{2}\,\mathrm{tr}[\mathbf{A}] \pm \mathrm{tr}[|\mathbf{A}|] = \frac{1}{2}(p_i - p_j) \pm \frac{1}{2}\,\mathrm{tr}[|\mathbf{A}|],\end{aligned}$$

so that we have by the assumption

$$\begin{aligned}\left|\mathrm{tr}[\Pi^{+}\mathbf{A}] - p_i\right| &= \left|-\frac{1}{2}(p_i + p_j) + \frac{1}{2}\,\mathrm{tr}[\mathbf{A}]\right| \\ &= \frac{1}{2}\left|\mathrm{tr}[|\mathbf{A}|] - (p_i + p_j)\right| \le \frac{1}{2}\delta,\end{aligned}$$

and similarly,

$$\left|\mathrm{tr}[\Pi^{-}\mathbf{A}] + p_j\right| \le \frac{1}{2}\delta.$$

Now, writing $\sigma_i = (\Phi_i(\omega_{i,j}))^{\otimes m}$ and $\sigma_j = (\Phi_j(\omega_{i,j}))^{\otimes m}$, we have obviously, $\mathrm{tr}[\Pi^{-}\sigma_i] \ge 0$, and on the other hand,

$$p_i\,\mathrm{tr}[\Pi^{-}\sigma_i] = \mathrm{tr}[\Pi^{-}\mathbf{A}] + p_j\,\mathrm{tr}[\Pi^{-}\sigma_j] \le -p_j + \frac{1}{2}\delta + p_j = \frac{1}{2}\delta.$$

The first result thus follows from $\Pi^{+} + \Pi^{-} = \mathbf{I}_m$ and $\mathrm{tr}[\sigma_i] = 1$. Similarly,

$$p_j\,\mathrm{tr}[\Pi^{+}\sigma_j] = -\,\mathrm{tr}[\Pi^{+}\mathbf{A}] + p_i\,\mathrm{tr}[\Pi^{+}\sigma_i] \le -p_i + \frac{1}{2}\delta + p_i = \frac{1}{2}\delta.$$

This proves the lemma. □

To compare the outputs of all the different branches of the channel, we define projections $\tilde{\Pi}_i$ on the tensor product space $\bigotimes_{1\le i<j\le M} \mathbb{H}_B^{\otimes m} = \mathbb{H}_B^{\otimes mL}$ with $L = \binom{M}{2}$ as follows:

$$\tilde{\Pi}_i = \bigotimes_{1\le i_1<i_2\le M} \Gamma^{(i)}_{i_1,i_2}, \tag{16.12}$$

where

$$\Gamma^{(i)}_{i_1,i_2} = \begin{cases} \mathbf{I}_m & \text{if } i_1 \neq i \text{ and } i_2 \neq i \\ \Pi^-_{i_1,i} & \text{if } i_2 = i \\ \Pi^+_{i,i_2} & \text{if } i_1 = i. \end{cases}$$

Notice that it follows from the fact that $\Pi^+_{i,j}\Pi^-_{i,j} = \mathbf{0}$, that the projections $\tilde{\Pi}_i$ are also disjoint:

$$\tilde{\Pi}_i\tilde{\Pi}_j = 0 \quad \text{if } i \neq j.$$

Introducing the notation,

$$\omega^{(mL)} = \bigotimes_{i_1<i_2} \omega^{\otimes m}_{i_1,i_2}, \tag{16.13}$$

we now have the following.

Lemma 16.2.4. *For all* $i = 1, \ldots, M$,

$$\lim_{m\to+\infty} \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] = 1. \tag{16.14}$$

Proof. Notice that for all $i < j$,

$$F(p_i\Phi_i(\omega_{i,j})^{\otimes m}, p_j\Phi_j(\omega_{i,j})^{\otimes m}) = \sqrt{p_ip_j}F(\Phi_i(\omega_{i,j}), \Phi_j(\omega_{i,j}))^m < f^m.$$

Using the inequalities,

$$\operatorname{tr}[\mathbf{A}_1] + \operatorname{tr}[\mathbf{A}_2] - 2F(\mathbf{A}_1, \mathbf{A}_2) \le \|\mathbf{A}_1 - \mathbf{A}_2\|_1 \le \operatorname{tr}[\mathbf{A}_1] + \operatorname{tr}[\mathbf{A}_2]$$

for any two positive operators $\mathbf{A}_1$ and $\mathbf{A}_2$, we find that

$$\left|\operatorname{tr}[|\mathbf{A}^{(m)}_{i,j}|] - (p_1 + p_2)\right| \le 2f^m,$$

since

$$\operatorname{tr}[|\mathbf{A}^{(m)}_{i,j}|] = \|p_i\Phi_i(\omega_{i,j})^{\otimes m} - p_j\Phi_j(\omega_{i,j})^{\otimes m}\|_1.$$

Using Lemma 16.2.3, we then have

$$
\begin{aligned}
1 &\geq \operatorname{tr}\Bigg[\tilde{\Pi}_i \Phi_i^{\otimes mL}\Bigg(\bigotimes_{i_1<i_2} \omega_{i,j}^{\otimes m}\Bigg)\Bigg] \\
&= \prod_{i_1 \leq i} \operatorname{tr}[\Pi_{i_1,i}^{-}(\Phi_i(\omega_{i_1,i})^{\otimes m})] \prod_{i_2 \geq i} \operatorname{tr}[\Pi_{i,i_2}^{+}(\Phi_i(\omega_{i,i_2})^{\otimes m})] \\
&\geq \left(1 - \frac{f^m}{p_i}\right)^{M-1}.
\end{aligned}
$$

This proves the lemma. □

We now fix m so large that

$$
\operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] > 1 - \delta \tag{16.15}
$$

for all $i = 1, 2, \ldots, M$.

The product state $\omega^{(mL)}$, defined through (16.13) is used as a preamble to the input state encoding each message, and serves to distinguish between the different branches, $\Phi_i, i = 1, 2, \ldots, M$, of the channel. If $\rho_k^{(n)} \in \mathfrak{B}(\mathbb{H}_A^{\otimes n})$ is a product state encoding the kth classical message in the set Λ_n, then the kth codeword is given by the product state

$$
\omega^{(mL)} \otimes \rho_k^{(n)}, \quad \forall k = 1, 2, \ldots, N_n. \tag{16.16}
$$

B. Direct part of Theorem 16.2.1

To prove the direct part of the classical capacity $C_{\text{prod}}^{\text{lm}}(\Phi)$ of long-term memory channel Φ in (16.4), i. e., the fact that a rate $R < C_{\text{prod}}^{\text{lm}}(\Phi)$ is achievable, we employ below the quantum analogue of Feinstein's fundamental lemma for the class of channels defined by (16.1).

Consider the n-use of the quantum memory channel $\Phi^{(n)} : \mathcal{S}(\mathbb{H}_A^{\otimes n}) \to \mathcal{S}(\mathbb{H}_B^{\otimes n})$,

$$
\Phi^{(n)}(\rho_A^{(n)}) = \sum_{i=1}^{M} p_i \Phi_i^{\otimes n}(\rho_A^{(n)}), \quad \forall \rho_A^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n}),
$$

that is a convex combination of M memoryless channels $\Phi_i, i = 1, 2, \ldots, M$, as described in (16.1). For any ensemble of states $(\{p_i\}_{i=1}^M, \{\rho_i\}_{i=1}^M)$ where $\rho_i \in \mathcal{S}(\mathbb{H}_A)$, define

$$
\hat{\chi}_\Phi(\{p_i, \rho_i\}) := \min\{\chi_i(\{p_i, \rho_i\}), i = 1, 2, \ldots, M\}, \tag{16.17}
$$

where $\chi_i(\{p_j, \rho_j\}) = \chi_{\Phi_i}(\{p_j, \rho_j\}) := \chi(\{p_j, \Phi_i(\rho_j)\})$, for $i = 1, \ldots, M$.

Theorem 16.2.5 (Feinstein's theorem for long-term memory channels). *Given $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ there exist at least $N_n \geq 2^{n(C_{\text{prod}}^{\text{lm}}(\Phi)-\epsilon)}$ product states $\rho_1^{(n)}, \ldots, \rho_{N_n}^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n})$ and positive operators $\mathbf{D}_1^{(n)}, \ldots, \mathbf{D}_{N_n}^{(n)} \in \mathfrak{B}(\mathbb{H}_B^{\otimes n})$ such that $\sum_{i=1}^{N_n} \mathbf{D}_i^{(n)} \leq \mathbf{I}_n$ and such that for each $k = 1, \ldots, N_n$,*

$$\mathrm{tr}[\Phi^{(n)}(\rho_k^{(n)})\mathbf{D}_k^{(n)}] > 1 - \epsilon. \tag{16.18}$$

Here,

$$C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi) := \sup_{\{p_j,\rho_j\}} \{\min\{\chi_i(\{p_j,\rho_j\}), i = 1,2,\dots,M\}\}, \tag{16.19}$$

where the supremum is taken over all finite ensemble of states ρ_j with probability p_j.

Remark 16.1. Note that the above theorem implies that a rate $R < C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)$ is achievable. This can be seen as follows: Given an $R < C_{\mathrm{prop}}^{\mathrm{lm}}(\Phi)$, choose $\epsilon > 0$ such that $R < C_{\mathrm{prop}}^{\mathrm{lm}}(\Phi) - \epsilon$. Then Theorem 16.2.5 guarantees the existence of codes $\mathfrak{C}^{(n)}$ of length n and size

$$N_n \geq 2^{n(C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)-\epsilon)} \geq 2^{nR},$$

with preamble-included codewords given by product states $\omega^{(mL)}\rho_j^{(n)}$, and POVM elements $\{\mathbf{D}_j^{(n)}\}$, for which the probability of error for the codes $\mathfrak{C}^{(n)}$, $\mathbb{P}_{\mathrm{err}}(\mathfrak{C}^{(n)})$ can be made arbitrarily small, for each $j \in \{1,2,\dots,N_n\}$ and n large enough. Hence, the rate R is achievable.

Proof of Theorem 16.2.5. Since (16.19) holds, for every $\epsilon > 0$ we can choose an ensemble $\{p_j,\rho_j\}_{j=1}^J$ such that

$$C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi) < \chi_i(\{p_j,\rho_j\}) + \frac{1}{4}\epsilon, \quad \forall i = 1,2,\dots,M. \tag{16.20}$$

Define $\sigma_{i,j} = \Phi_i(\rho_j)$, $\sigma_{i,\vec{j}}^{(n)} = \otimes_{r=1}^n \sigma_{i,j_r}$, $\bar{\sigma}_i = \sum_{j=1}^J p_j\Phi_i(\rho_j) = \Phi_i(\bar{\rho})$, and $\bar{\sigma}_i^{(n)} = \bar{\sigma}_i^{\otimes n}$, where $\vec{j} = (j_1,j_2,\dots,j_n)$. Let $\bar{\mathbf{P}}_i^{(n)}$, $i = 1,2,\dots,M$ be the orthogonal projections on the typical subspaces for the states $\bar{\sigma}_i^{(n)}$. Then, by Lemma 13.2.4, we have

$$\bar{\mathbf{P}}_i^{(n)}(\bar{\sigma}_i) > 1 - \delta^2 \tag{16.21}$$

for n large enough, and

$$\bar{\mathbf{P}}_i^{(n)}\bar{\sigma}_i^{(n)}\bar{\mathbf{P}}_i^{(n)} \leq 2^{-n(H(\bar{\sigma}_i)-\frac{1}{\epsilon})}. \tag{16.22}$$

By Lemma 13.2.6, there also exist typical subspaces with projections $\mathbf{P}_{i,\vec{j}}^{(n)}$ for which

$$\mathbb{E}(\mathrm{tr}[\sigma_{i,\vec{j}}^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}]) > 1 - \delta^2 \tag{16.23}$$

for sufficiently large n. Let $N = \tilde{N}(n)$ be the maximal number of product states $\tilde{\rho}_1^{(n)},\dots,\tilde{\rho}_N^{(n)}$ on $\mathbb{H}_A^{\otimes n}$ (each of which is a tensor product of states in the maximizing ensemble $\{p_j,\rho_j\}_{j=1}^J$) for which there exist positive operators $\mathbf{D}_1^{(n)},\dots,\mathbf{D}_N^{(n)}$ on $\mathbb{H}_B^{\otimes mL} \otimes \mathbb{H}_B^{\otimes n}$ such that:

1. $\mathbf{D}_k^{(n)} = \sum_{i=1}^M \tilde{\Pi}_i \otimes \mathbf{D}_{k,i}^{(n)}$ and $\sum_{k=1}^N \mathbf{D}_{k,i}^{(n)} \le \bar{\mathbf{P}}_i^{(n)}$;
2. $\sum_{i=1}^M p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \operatorname{tr}[\Phi_i(\tilde{\rho}_k^{(n)})\mathbf{D}_{k,i}^{(n)}] > 1-\epsilon$; and
3. $\sum_{i=1}^M p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \operatorname{tr}[(\Phi_i(\bar{\rho}))^{\otimes n}\mathbf{D}_{k,i}^{(n)}] \le 2^{-n(C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)-\frac{1}{2}\epsilon)}$,

for $\bar{\rho} = \sum_{j=1}^J p_j \rho_j$. For each $i = 1,\dots,M$ and $\vec{j} = (j_1,\dots,j_n) \in J^{\times n}$, we define, as before

$$\mathbf{V}_{i,\vec{j}}^{(n)} = \left(\bar{\mathbf{P}}_i^{(n)} - \sum_{k=1}^N \mathbf{D}_{k,i}^{(n)}\right)^{1/2} \bar{\mathbf{P}}_i^{(n)} \mathbf{P}_{i,\vec{j}}^{(n)} \bar{\mathbf{P}}_i^{(n)} \left(\bar{\mathbf{P}}_i^{(n)} - \sum_{k=1}^N \mathbf{D}_{k,i}^{(n)}\right)^{1/2} \tag{16.24}$$

Clearly, $\mathbf{V}_{i,\vec{j}}^{(n)} \le \bar{\mathbf{P}}_i^{(n)} - \sum_{k=1}^N \mathbf{D}_{k,i}^{(n)}$. Put

$$\mathbf{V}_{\vec{j}}^{(n)} = \sum_{i=1}^M \tilde{\Pi}_i \otimes \mathbf{V}_{i,\vec{j}}^{(n)}. \tag{16.25}$$

This is a candidate for an additional measurement operator, $\mathbf{D}_{N+1}^{(n)}$, for Bob with corresponding input state $\tilde{\rho}_{N+1}^{(n)} = \rho_{\vec{j}}^{(n)} = \rho_{j_1} \otimes \rho_{j_2} \otimes \cdots \rho_{j_n}$. Clearly, the condition (1) given above is satisfied and we also have the following.

Lemma 16.2.6. *The following inequality holds:*

$$\sum_{i=1}^M p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \operatorname{tr}[\bar{\sigma}_i^{(n)} \mathbf{V}_{i,\vec{j}}^{(n)}] \le 2^{-n(C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)-\frac{1}{2}\epsilon)}, \tag{16.26}$$

where $\bar{\sigma}_i^{(n)} = (\Phi_i(\rho))^{\otimes n}$.

Proof. By the definition of the typical subspaces, we have

$$\operatorname{tr}[\bar{\sigma}^{(n)} \mathbf{V}_{\vec{j}}^{(n)}] \le 2^{-n[H(\bar{\sigma}_i)-\bar{H}-\frac{1}{2}\epsilon]} = 2^{-n[\chi_i-\frac{1}{2}\epsilon]}, \tag{16.27}$$

for n large enough, where $\chi_i(\{p_j,\rho_j\}) = \chi_{\Phi_i}(\{p_j,\rho_j\}) := \chi(\{p_j,\Phi_i(\rho_j)\})$, for $i = 1,\dots,M$. Then

$$\begin{aligned}
&\sum_{i=1}^M p_i \operatorname{tr}[\tilde{\Pi}_i^- \Phi_i^{\otimes mL}(\omega^{(mL)})] \operatorname{tr}[\bar{\sigma}_i^{(n)} \mathbf{V}_{i,\vec{j}}^{(n)}] \\
&\quad \le \operatorname{tr}[\bar{\sigma}_i^{(n)} \mathbf{V}_{i,\vec{j}}^{(n)}] \le \sum_{i=1}^M p_i 2^{-n[H(\bar{\sigma}_i)-\bar{H}-\frac{1}{2}\epsilon]} \\
&\quad \le 2^{-n[\hat{\chi}(\Phi)-\frac{1}{2}\epsilon]} \le 2^{-n[C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)-\frac{1}{2}\epsilon]},
\end{aligned} \tag{16.28}$$

where we used the fact that $\operatorname{tr}[\tilde{\Pi}_i^- \Phi_i^{\otimes mL}(\omega^{(mL)})] \le 1$ and $\hat{\chi}_\Phi(\{p_i,\rho_i\}) := \min_{i=1,\dots,M} \chi_i(\{p_i,\rho_i\})$. This proves the lemma. ☐

By maximality of N, it now follows that the condition (2) above cannot hold, i. e.,

$$\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \operatorname{tr}[\Phi_i^{\otimes n}(\rho_{\vec{j}}^{(n)})\mathbf{V}_{i,\vec{j}}^{(n)}] \le 1-\epsilon \tag{16.29}$$

for every $\mathbf{j}$, and this yields the following.

Corollary 16.2.7.

$$\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \operatorname{tr}[\Phi_i^{\otimes n}(\rho_{\mathbf{j}}^{(n)})\mathbf{V}_{i,\vec{j}}^{(n)}] \le 1-\epsilon \tag{16.30}$$

We also need the following lemma.

Lemma 16.2.8. *For all $\eta' \ge \delta^2 + 3\delta$,*

$$\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \operatorname{tr}[\sigma_{i,\vec{j}}^{(n)} \bar{\mathbf{P}}_i^{(n)} \mathbf{P}_{i,\vec{j}}^{(n)} \bar{\mathbf{P}}_i^{(n)}] \le 1-\eta', \tag{16.31}$$

if n is large enough.

Proof. Using Corollary 16.2.7 and 16.15, we have

$$\begin{aligned}\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)} \bar{\mathbf{P}}_i^{(n)} \mathbf{P}_{i,\vec{j}}^{(n)} \bar{\mathbf{P}}_i^{(n)}]) \\ \le (1-\eta)(1-\delta),\end{aligned} \tag{16.32}$$

provided $\eta \ge \delta^2 + 2\delta$. Hence, the results follows. This proves the lemma. □

Lemma 16.2.9. *Assume $\eta' < \frac{1}{3}\epsilon$ and write*

$$\mathbf{Q}_i^{(n)} = \sum_{k=1}^{N} \mathbf{D}_{k,i}^{(n)} \tag{16.33}$$

Then for n large enough,

$$\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \mathbb{E}(\operatorname{tr}[\Phi_i^{(n)}(\rho_{\vec{j}}^{(n)})\mathbf{Q}_i^{(n)}]) \ge \eta'^2. \tag{16.34}$$

Proof. Define

$$\mathbf{R}_i^{(n)} = \bar{\mathbf{P}}^{(n)} - \sqrt{\bar{\mathbf{P}}_i^{(n)} - \mathbf{Q}_i^{(n)}}.$$

By the Corollary 16.2.7,

$$
\begin{aligned}
1-\epsilon &\geq \sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})]\mathbb{E}(\operatorname{tr}[\Phi_i^{(n)}(\rho_{\vec{j}}^{(n)})\mathbf{V}_{i,\vec{j}}^{(n)}]) \\
&= \sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})]\mathbb{E}(\operatorname{tr}[\Phi_i^{(n)}(\rho_{\vec{j}}^{(n)})\mathbf{Q}_i^{(n)}]) \\
&\quad - \sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \\
&\quad \times \mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}\mathbf{R}_i^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}\bar{\mathbf{P}}_i^{(n)}] + \operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}\bar{\mathbf{P}}_i^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}\mathbf{R}_i^{(n)}]) \\
&\quad + \sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})]\mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}\mathbf{R}_i^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}\mathbf{R}_i^{(n)}]).
\end{aligned}
$$

Since the last term is positive, we have by Lemma 16.2.8,

$$
\begin{aligned}
&\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \\
&\quad \times \mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}\mathbf{R}_i^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}\bar{\mathbf{P}}_i^{(n)}] + \operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}\bar{\mathbf{P}}_i^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}\mathbf{R}_i^{(n)}]) \geq \epsilon - 2\eta' > 2\eta'.
\end{aligned}
$$

On the other hand, using the Cauchy–Schwarz inequality for each term, we have

$$
\begin{aligned}
&\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})] \\
&\qquad \times \mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}\mathbf{R}_i^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}\bar{\mathbf{P}}_i^{(n)}] + \operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}\bar{\mathbf{P}}_i^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}\mathbf{R}_i^{(n)}]) \\
&\quad \leq 2\left\{\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})]\mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}(\mathbf{R}_i^{(n)})^2])\right\}^{1/2} \\
&\qquad \times \left\{\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})]\mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}\bar{\mathbf{P}}_i^{(n)}\mathbf{P}_{i,\vec{j}}^{(n)}\bar{\mathbf{P}}_i^{(n)}])\right\}^{1/2} \\
&\quad \leq 2\left\{\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})]\mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}(\mathbf{R}_i^{(n)})^2)]\right\}^{1/2}.
\end{aligned}
$$

Thus,

$$
\sum_{i=1}^{M} p_i \operatorname{tr}[\tilde{\Pi}_i \Phi_i^{\otimes mL}(\omega^{(mL)})]\mathbb{E}(\operatorname{tr}[\sigma_{i,\vec{j}}^{(n)}(\mathbf{R}_i^{(n)})^2]) \geq \eta'^2.
$$

To complete the proof, we remark as before that

$$
\mathbf{Q}_i^{(n)} \geq (\mathbf{R}_i^{(n)})^2.
$$

This proves the lemma. □

To complete the proof of Theorem 16.2.5, we note as before for large enough n that $\tilde{N}(n) \geq (\eta')^2 2^{n[C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi)-\frac{3}{4}\epsilon]}$. We take the following states as codewords:

$$\rho_k^{(mL+n)} = \omega^{(mL)} \otimes \tilde{\rho}_k^{(n)}, \tag{16.35}$$

where $\omega^{(mL)}$ is the preamble defined by (16.13). Consequently, for sufficiently large n, we have

$$N_{mL+n} = \tilde{N}(n) \geq (\eta')^2 2^{n[C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi)-\frac{3}{4}\epsilon]} \geq 2^{(mL+n)(C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi)-\epsilon)}.$$

To complete the proof, we need to show that the set $\{\mathbf{D}_k^{(n)}\}$ satisfies (16.18). But this follows immediately from condition (2) stated earlier:

$$\begin{aligned}
&\mathrm{tr}[\Phi^{(mL+n)}(\rho_k^{(mL+n)})\mathbf{D}_k^{(n)}] \\
&= \sum_{i=1}^{M} p_i \,\mathrm{tr}[\Phi_i^{\otimes mL+n}(\omega^{(mL)} \otimes \tilde{\rho}_k^{(n)})\mathbf{D}_k^{(n)}] \\
&= \sum_{i,j=1}^{M} p_i \,\mathrm{tr}[\tilde{\Pi}_j \Phi^{\otimes mL}(\omega^{(mL)})]\,\mathrm{tr}[\Phi_i^{\otimes n}(\tilde{\rho}_k^{(n)})\mathbf{D}_{k,j}^{(n)}] \\
&\geq \sum_{i,j=1}^{M} p_i \,\mathrm{tr}[\tilde{\Pi}_i \Phi^{\otimes mL}(\omega^{(mL)})]\,\mathrm{tr}[\Phi_i^{\otimes n}(\tilde{\rho}_k^{(n)})\mathbf{D}_{k,i}^{(n)}] \\
&\geq 1 - \epsilon.
\end{aligned}$$

This proves Theorem 16.2.5. □

C. Weak converse of Theorem 16.2.1

We first recall Fano's inequality without proof as follows (see Fano [49]). Suppose the random variables X and Y represent input and output messages with a joint probability $p(x, y)$. Let the event $\{X \neq \tilde{X}\}$ be the event representing an occurrence of error with $\tilde{X} = f(Y)$ being an approximate version of X. Let $P_{\mathrm{err}} := \Pr\{X \neq \tilde{X}\}$ be the probability of error. Then we have the following Fano's inequality holds:

$$H(X|Y) \leq h(P_{\mathrm{err}}) + P_{\mathrm{err}} \log(|\mathcal{X}| - 1), \tag{16.36}$$

where $\mathcal{X}$ denotes the support of X and

$$H(X|Y) = -\sum_{i,j} p(x_i, y_j) \log p(x_i|y_j) \tag{16.37}$$

is the conditional entropy and

$$h_2(P_{\mathrm{err}}) = -P_{\mathrm{err}} \log P_{\mathrm{err}} - (1 - P_{\mathrm{err}}) \log(1 - P_{\mathrm{err}})$$

is the corresponding binary entropy.

Fano's inequality is often used to find a lower bound on the error probability of any decoder as well as the lower bounds for minimax risks in density estimation in information theory.

Going back to providing a proof of weak converse of Theorem 16.2.1, we prove that it is impossible for Alice to transmit classical messages reliably to Bob through the long term memory channel Φ defined in (16.1) at a rate $R > C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi)$. This is the weak converse of Theorem 16.2.1 in the sense that the probability of error does not tend to zero asymptotically as the length of the code increases, for any code with rate $R > C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi)$. To prove the weak converse, suppose that Alice encodes messages labeled by $\lambda \in \Lambda_n$ by product states $\rho^{(n)}_\lambda = \rho_{\lambda,1} \otimes \cdots \otimes \rho_{\lambda,n}$ in $\mathcal{S}(\mathbb{H}^{\otimes n}_A)$. Let the corresponding outputs for the ith branch of the channel be denoted by $\sigma^{(n)}_{\lambda,i}$, i. e.,

$$\sigma^{(n)}_{\lambda,i} = \Phi^{\otimes n}_i(\rho^{(n)}_\lambda) = \sigma^i_{\lambda,1} \otimes \cdots \otimes \sigma^i_{\lambda,n}, \quad \sigma^i_{\lambda,j} = \Phi_i(\rho_{\lambda,j}). \tag{16.38}$$

We further define

$$\bar{\sigma}^{(n)}_{\lambda,i} = \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \sigma^{(n)}_{\lambda,i} \tag{16.39}$$

and

$$\bar{\sigma}_{i,j} = \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \sigma^i_{\lambda,j}. \tag{16.40}$$

Let Bob's POVM elements corresponding to the codewords $\rho^{(n)}_\lambda$ be denoted by $\mathbf{D}^{(n)}_\lambda$, $\lambda \in \Lambda_n$. We may assume that Alice's messages are produced uniformly at random from the set Λ_n. Then Bob's average probability of error is given by

$$\bar{\mathbb{P}}^{(n)}_{\mathrm{err}} := 1 - \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \mathrm{tr}[\Phi^{(n)}(\rho^{(n)}_\lambda)\mathbf{D}^{(n)}_\lambda]. \tag{16.41}$$

We also define the average error corresponding to the ith branch of the channel

$$\bar{\mathbb{P}}^{(n)}_{i,\mathrm{err}} = 1 - \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \mathrm{tr}[\Phi^{\otimes n}_i(\rho^{(n)}_\lambda)\mathbf{D}^{(n)}_\lambda]. \tag{16.42}$$

So, that

$$\bar{\mathbb{P}}^{(n)}_{\mathrm{err}} = \sum_{i=1}^{M} p_i \bar{\mathbb{P}}^{(n)}_{i,\mathrm{err}}. \tag{16.43}$$

Let $X^{(n)}$ be a random variable with a uniform distribution over the set Λ_n, characterizing the classical message sent by Alice to Bob. Let $Y^{(n)}_i$ be the random variable corresponding to Bob's inference of Alice's message, when the codeword $\rho^{(n)}_i$ is transmitted through the ith branch, $\Phi_i^{\otimes n}$ of the channel $\Phi^{(n)}$. It is defined by the conditional probabilities

$$\Pr[Y_i^{(n)} = y | X^{(n)} = \lambda] = \operatorname{tr}[\Phi_i^{\otimes n}(\rho_\lambda^{(n)})\mathbf{D}_y^{(n)}]. \tag{16.44}$$

By Fano's inequality (see (16.36)),

$$\begin{aligned} &h_2(\bar{\mathbb{P}}_{i,\mathrm{err}}^{(n)}) + \bar{\mathbb{P}}_{i,\mathrm{err}}^{(n)} \log(|\Lambda_n| - 1) \\ &\quad \geq H(X^{(n)}|Y_i^{(n)}) = H(X^{(n)}) - H(X^{(n)} \| Y_i^{(n)}) \end{aligned} \tag{16.45}$$

Here, $h_2(\cdot)$ denotes the binary entropy and $H(\cdot)$ denotes the Shannon entropy. Using the Holevo bound and the subadditivity of the von Neumann entropy (see Proposition 7.2.8), we have

$$\begin{aligned} H(X^{(n)} \| Y_i^{(n)}) &\leq H\bigg(\frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} \Phi_i^{\otimes n}(\rho_\lambda^{(n)})\bigg) - \frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} H(\Phi_i^{\otimes n}(\rho_\lambda^{(n)})) \\ &= H\bigg(\frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} \sigma_{\lambda,i}^{(n)}\bigg) - \frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} H(\sigma_{\lambda,i}^{(n)}) \\ &\leq \sum_{j=1}^{n}\bigg[H(\bar{\sigma}_{i,j}^{(n)}) - \frac{1}{|\Lambda_n|}\sum_{\lambda\in\Lambda_n} H(\sigma_{\lambda,j}^{i})\bigg] \\ &= \sum_{j=1}^{n}\chi_i\bigg(\bigg\{\frac{1}{|\Lambda_n|}, \rho_{\lambda,j}\bigg\}_{\lambda\in\Lambda_n}\bigg) \\ &= \sum_{j=1}^{n}\frac{1}{|\Lambda_n|} H(\sigma_{\lambda,j}^{i} \| \bar{\sigma}_{i,j}) \end{aligned} \tag{16.46}$$

The latter expression can be rewritten using Donald's identity restated below:

$$\sum_j p_j H(\omega_j \| \rho) = \sum_j p_j H(\omega_j \| \bar{\omega}) + H(\bar{\omega} \| \rho), \tag{16.47}$$

where $\bar{\omega} = \sum_j p_j \omega_j$. We apply this with ρ replaced by

$$\bar{\sigma}_i = \frac{1}{n|\Lambda_n|}\sum_{j=1}^{n}\sum_{\lambda\in\Lambda_n} \sigma_{\lambda,j}^{i} \tag{16.48}$$

and the sum replaced by a double sum over j and λ with states $\sigma_{\lambda,j}^{i}$. This yields

$$\begin{aligned} &\frac{1}{n|\Lambda_n|}\sum_{j=1}^{n}\sum_{\lambda\in\Lambda_n} H(\sigma_{\lambda,j}^{i} \| \bar{\sigma}_{i,j}) \\ &\quad = \frac{1}{n|\Lambda_n|}\sum_{j=1}^{n}\sum_{\lambda\in\Lambda_n} H(\sigma_{\lambda,j}^{i} \| \bar{\sigma}_i) + H(\bar{\sigma}_i \| \bar{\sigma}_{i,j}) \end{aligned} \tag{16.49}$$

But it follows from the convexity of relative entropy that the right-hand side of the second term equals zero:

$$0 \le H(\bar{\sigma}_i \| \bar{\sigma}_{i,j}) \le \frac{1}{n}\sum_{j=1}^{n} H(\bar{\sigma}_{i,j} \| \bar{\sigma}_{i,j}) = 0. \tag{16.50}$$

Inserting into (16.46), we now have

$$\begin{aligned}\frac{1}{n}H(X^{(n)} \| Y_i^{(n)}) &\le \frac{1}{n|\Lambda_n|}\sum_{j=1}^{n}\sum_{\lambda\in\Lambda_n} H(\sigma^i_{\lambda,j} \| \bar{\sigma}_i)\\ &= \chi_i\left(\left\{\frac{1}{n|\Lambda_n|}, \rho_{\lambda,j}\right\}_{\lambda,j}\right)\end{aligned} \tag{16.51}$$

Fano's inequality (16.45) now yields

$$\begin{aligned}&h(\bar{\mathbb{P}}^{(n)}_{i,\mathrm{err}}) + \bar{\mathbb{P}}^{(n)}_{i,\mathrm{err}} \log(|\Lambda_n|)\\ &\quad\ge \log|\Lambda_n| - n\chi_i\left(\left\{\frac{1}{n|\Lambda_n|}, \rho_{\lambda,j}\right\}_{\lambda,j}\right).\end{aligned} \tag{16.52}$$

However, since

$$C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi) \ge \min\left\{\chi_i\left(\left\{\frac{1}{n|\Lambda_n}, \rho_{\lambda,j}\right\}_{\lambda,j}\right), i = 1, 2, \ldots, M\right\} \tag{16.53}$$

and $R = \frac{1}{n}\log|\Lambda_n| > C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi)$, there must be at least one branch i such that

$$\bar{\mathbb{P}}^{(n)}_{i,\mathrm{err}} \ge 1 - \frac{C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi) + 1/n}{R} > 0. \tag{16.54}$$

We conclude from (16.43) and (16.54) that

$$\bar{\mathbb{P}}^{(n)}_{\mathrm{err}} \ge \left(1 - \frac{C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi) + 1/n}{R}\right)\min\{p_i, i = 1, 2, \ldots, M\}. \tag{16.55}$$

This proves the theorem. □

16.2.2 Constrained classical capacity in infinite dimensions

Let $\mathbf{H}$ be a positive self-adjoint operator on the input system $\mathbb{H}_A$ and consider the linearly constrained set $\mathcal{A} = \mathcal{K}_{\mathbf{H}}(E) \subset \mathcal{S}(\mathbb{H}_A)$ described by

$$\mathcal{K}_{\mathbf{H}}(E) := \{\rho \in \mathcal{S}(\mathbb{H}_A) \mid \mathrm{tr}[\rho\mathbf{H}] \le E\}, \quad E \ge 0. \tag{16.56}$$

For arbitrary state $\rho \in \mathcal{S}(\mathbb{H}_A)$ with spectral decomposition $\rho = \sum_i \lambda_i |\phi_i\rangle_A \langle\phi_i|$, we define

$$\mathrm{tr}[\rho\mathbf{H}] := \sum_i \lambda_i \|\sqrt{\mathbf{H}}|\phi_i\rangle\|^2_{\mathbb{H}_A} \leq +\infty.$$

As noted earlier the n-use of a memoryless quantum channel can be written as $\Phi^{(n)} = \Phi^{\otimes n}$, where $\Phi^{\otimes n}$ is the tensor product channel on the Hilbert space $\mathbb{H}_A^{\otimes n}$. In this case, the observable $\mathbf{H}^{(n)}$ on $\mathbb{H}_A^{\otimes n}$ corresponding to $\mathbf{H}$ on $\mathbb{H}_A$ can be defined as

$$\mathbf{H}^{(n)} = \mathbf{H} \otimes \mathbf{I}_A \otimes \cdots \otimes \mathbf{I}_A + \cdots + \mathbf{I}_A \otimes \cdots \otimes \mathbf{I}_A \otimes \mathbf{H}, \tag{16.57}$$

where $\mathbf{I}_A$ is the identity operator on the input system $\mathbb{H}_A$. We want the input state $\rho^{(n)}$ on the tensor product space $\mathbb{H}_A^{\otimes n}$ to satisfy the additive constraint

$$\mathrm{tr}[\rho^{(n)}\mathbf{H}^{(n)}] \leq nE. \tag{16.58}$$

We make the following basic assumption on energy constraint on each branch $\Phi_i; \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$, $i = 1, 2, \ldots, M$, of memoryless channels.

Assumption 16.1. The following constraint on each of the memoryless channel Φ_i, $i = 1, 2, \ldots, M$, is imposed:

$$\sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} H(\Phi_i(\rho)) < +\infty, \quad \forall i = 1, 2, \ldots, M, \tag{16.59}$$

where E is a positive constant and $H(\cdot) : \mathcal{S}(\mathbb{H}_A) \to [0, +\infty]$ is the von Neumann entropy.

Lemma 16.2.10. *Let Φ be the channel with long-term memory defined by* (16.1). *Then under Assumption* 16.1*,*

$$\sup_{\rho^{(n)} \in \mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} H(\Phi^{\otimes n}(\rho^{(n)})) < +\infty, \tag{16.60}$$

where

$$\mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE) = \{\rho^{(n)} \in \mathcal{S}(\mathbb{H}_A^{\otimes n}) \mid \mathrm{tr}[\rho^{(n)}\mathbf{H}^{(n)}] \leq nE\}. \tag{16.61}$$

Proof. Based on Assumption 16.1, we have the following immediately result on the channel $\Phi(\rho) = \sum_{i=1}^M p_i \Phi_i(\rho)$,

$$\sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} H(\Phi(\rho)) < +\infty. \tag{16.62}$$

This is because

$$H(\Phi(\rho)) = H\left(\sum_{i=1}^{M} p_i \Phi_i(\rho)\right) \le \sum_{i=1}^{M} p_i H(\Phi_i(\rho)) < +\infty \tag{16.63}$$

by the convexity of the von Neumann entropy function $H(\cdot)$ (see Theorem 8.1.6). Note that (13.38) implies similar property of the channel $\Phi_i^{\otimes n}$;

$$\sup_{\rho^{(n)} \in \mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} H(\Phi_i^{\otimes n}(\rho^{(n)})) < +\infty. \tag{16.64}$$

Indeed, the subadditivity of von Neumann entropy with respect to tensor product

$$H(\Phi_i^{\otimes n}(\rho^{(n)})) \le \sum_{k=1}^{n} H(\Phi_i(\rho_k^{(n)})),$$

where $\rho_k^{(n)}$ is the kth partial state of $\rho^{(n)}$. Also, by concavity of the entropy

$$\sum_{k=1}^{n} H(\Phi_i(\rho_k^{(n)})) \le nH(\Phi_i(\bar{\rho}^{(n)})),$$

where $\bar{\rho}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} \rho_k^{(n)}$. The inequality (16.58) can be rewritten as

$$\frac{1}{n}\sum_{k=1}^{n} \operatorname{tr}[\rho_k^{(n)} \mathbf{H}^{(\mathbf{n})}] = \operatorname{tr}[\bar{\rho}^{(n)} \mathbf{H}^{(n)}] \le E,$$

which implies that

$$\sup_{\rho^{(n)} \in \mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} H(\Phi_i^{\otimes n}(\rho^{(n)})) \le n \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} H(\Phi_i(\rho)).$$

Consequently,

$$\begin{aligned}
\sup_{\rho^{(n)} \in \mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} H(\Phi^{(n)}(\rho^{(n)})) &= \sup_{\rho^{(n)} \in \mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} H\left(\sum_{i=1}^{M} p_i \Phi_i^{\otimes n}(\rho^{(n)})\right) \\
&= \sup_{\rho^{(n)} \in \mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} \sum_{i=1}^{M} p_i H(\Phi_i^{\otimes n}(\rho^{(n)})) \\
&\le n \sup_{\rho^{(n)} \in \mathcal{K}^{(n)}_{\mathbf{H}^{(n)}}(nE)} \sum_{i=1}^{M} p_i H(\Phi_i(\rho)) \\
&< +\infty.
\end{aligned} \tag{16.65}$$

This proves the lemma. □

In this subsection, we investigate classical capacity $C^{\mathrm{lm}}_{\mathrm{prod}}(\Phi; \mathbf{H}, E)$ with each of Φ_i, $i = 1, 2, \ldots, M$ that satisfying Assumption 16.1.

For infinite-dimensional memoryless channel Φ satisfying the constraint (16.1), the code, error probability and classical capacity for the channel is defied below.

Definition 16.2.11. Let $\Phi^{(n)} : \mathcal{S}(\mathbb{H}_A^{\otimes n}) \to \mathcal{S}(\mathbb{H}_B^{\otimes n})$ be the n-use of the channel Φ defined by (16.1).

1. The triplet $\mathfrak{C}^{(n)} = (p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)})$ is said to be a code of length n and of size N_n, where (i) $p^{(n)} = \{p_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a probability distribution, i. e., $p_j^{(n)} > 0$ for all $j = 1, 2, \ldots, N_n$ with $\sum_{j=1}^{N_n} p_j^{(n)} = 1$; (ii) $\rho^{(n)} = \{\rho_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a collection of N_n states satisfying (16.58) and (iii) $\mathbf{D}^{(n)} = \{\mathbf{D}_j^{(n)} \mid j = 1, 2, \ldots, N_n\}$ is a POVM on $\mathbb{H}_A^{\otimes n}$ that represents the decoding operators used by the receiver, Bob.
2. The average error probability $\bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)})$ for the code $\mathfrak{C}^{(n)}$ is defined by
$$\bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)}) = \frac{1}{N_n} \sum_{j=1}^{N_n} p_j^{(n)} \{1 - \mathrm{tr}[\Phi^{(n)}(\rho_j^{(n)}) \mathbf{D}_j^{(n)}]\}.$$
Let $\bar{\mathbb{P}}_{\mathrm{err}}(n, N_n)$ be the average error probability for any code $\mathfrak{C}^{(n)}$ with length n and size N_n.
3. The energy constrained classical capacity $C_{\mathrm{prop}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E)$ of the channel Φ defined by (16.1) is defined as the least upper bound of the rates R for which $\liminf_{n\to+\infty} \mathbb{P}_{\mathrm{err}}(n, 2^{nR}) = 0$, i. e.,
$$C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E) = \inf\Big\{R > 0 \mid \liminf_{n\to+\infty} \bar{\mathbb{P}}_{\mathrm{err}}(n, 2^{nR}) = 0\Big\}, \tag{16.66}$$

The following result enables a computation of energy constrained $C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E)$. We omit the proof here because it is a special case of entanglement-assisted classical capacity $C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E)$, which is the center of investigation in the next section.

Theorem 16.2.12. *Let* $\mathbf{H}$ *be an* $\mathfrak{H}$*-operator defined on* $\mathbb{H}_A$*. Then the constrained classical capacity* $C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E)$ *of* Φ *defined by* (16.1) *satisfies the following equation:*

$$C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E) = \sup_{\rho^{(n)} \in \mathcal{K}_{\mathbf{H}^{(n)}}^{(n)}(nE)} \{\min\{\chi(\{p^{(n)}, \Phi_i(\rho^{(n)})\}), i = 1, 2, \ldots, M\}\}, \tag{16.67}$$

where $\chi(\{(p^{(n)}, \Phi_i(\rho^{(n)})\})$ *is the* χ*-quantity at the ensemble* $\{(p^{(n)}, \Phi_i(\rho^{(n)}))\}$ *defined in Definition* 12.1.1.

16.3 Entanglement-assisted classical capacity

Following the protocol for entanglement assisted communication for memoryless quantum channel introduced in Section 13.4, we follow the approaches used in Datta and Dorlas [33, 34], Datta–Suhov–Dorlas [35] and Holevo and Shirokov [83] to explore the entanglement assisted classical capacity for quantum channel with long term memory as follows.

Let A' be a quantum system which may be different from system B but is accessible by both Alice (system A) and Bob (system B). As defined in (16.1), let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be a channel with long term memory with

$$\Phi(\rho_A) = \sum_{i=1}^{M} p_i \Phi_i(\rho_A), \quad \forall \rho_A \in \mathcal{S}(\mathbb{H}_A),$$

where $\Phi_i : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$, $i = 1, 2, \ldots, M$ is a memoryless channel, and $p_i > 0$ with $\sum_{i=1}^{M} p_i = 1$.

Suppose Alice has a set of messages, labeled by the elements of the set $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{M_n}\}$, which she would like to communicate via the quantum channel (16.1) to Bob, exploiting this shared entanglement. In the entanglement-assisted channel as described by Holevo and Shrikov [83], Alice (quantum system A) and Bob (quantum system B) share indefinitely many copies of an entangled (pure) state ω_{AB}. Alice first makes encodings $\lambda \to \mathfrak{E}_A^\lambda$ of a classical signal λ, which was chosen at random from a finite alphabet Λ with probabilities $p_\lambda > 0$ (where $\sum_{\lambda \in \Lambda_n} p_\lambda = 1$) and then sends part of this shared state ω_{AB}^λ through the channel Φ to B. Here $\mathfrak{E}_A^\lambda$ are encoding channels depending on the signal λ. Thus, B receives states $(\Phi \otimes \mathfrak{I}_B)(\omega_{AB}^\lambda)$, where $\omega_{AB}^\lambda = (\mathfrak{E}_A^\lambda \otimes \mathfrak{I}_B)(\omega_{AB})$, with probabilities p_λ, and B aims to extract maximum information about λ by doing measurements on these states. To enable block encoding, this procedure should be applied to the channel $\Phi^{(n)}$. Then signal states $\omega_{AB}^{\lambda(n)}$ transmitted through the channel $\Phi^{(n)} \otimes \mathfrak{I}_B^{\otimes n}$ have a special form

$$\omega_{AB}^{\lambda(n)} = (\mathfrak{E}_A^{\lambda(n)} \otimes \mathfrak{I}_B^{(n)})(\omega_{AB}^{(n)}), \tag{16.68}$$

where $\omega_{AB}^{(n)}$ is the pure entangled state for n copies of the system AB and $\lambda \mapsto \mathfrak{E}_A^{\lambda(n)}$ are encodings of n copies of system A.

For this purpose, she uses encoding (CPTP) maps $\{\mathfrak{E}_A^\lambda\}_{\lambda \in \Lambda_n}$ acting on $\mathfrak{B}(\mathbb{H}_A)$. In order to transmit her classical messages through the quantum channel, Alice encodes each of her messages in a quantum state in $\mathbb{H}_{AB}^{\otimes n}$ in the following manner. To each $\lambda \in \Lambda_n$, she assigns a quantum state (or codeword) where

$$\omega_{AB}^{\lambda(n)} := \omega_\lambda^1 \otimes \cdots \otimes \omega_\lambda^n \in \mathfrak{B}(\mathbb{H}_{AB}^{\otimes n}), \tag{16.69}$$

where $\omega_k^\lambda = (\mathfrak{E}_{j_k} \otimes \mathfrak{I}_B)\psi^{AB} \in \mathfrak{B}(\mathbb{H}_{AB})$, for $k = 1, \ldots, n$. Here, $j_k \in \{1, \ldots, J\}$ and $\mathfrak{I}_B$ denotes the identity map in $\mathfrak{B}(\mathbb{H}_B)$. Note that the codewords are states shared between Alice and Bob. Alice then sends her part of these shared states to Bob through n subsequent uses of the quantum channel (16.1). Hence, Bob's final state corresponding to Alice's classical message λ is

$$\sigma_{AB}^{\lambda(n)} := (\Phi^{(n)} \otimes \mathfrak{I}_B^{\otimes n})(\omega_{AB}^{\lambda(n)}).$$

In order to infer the message that Alice communicated to him, Bob makes a measurement on the state $\sigma_{AB}^{\lambda(n)}$ he received, the measurement being described by POVM elements $\mathbf{D}_{AB}^{\lambda(n)}$, $\lambda \in \Lambda_n$, with $\mathbf{D}_{AB}^{\lambda,(n)}$ being a positive operator acting on $\mathbb{H}_{AB}^{\otimes n}$, such that

$$\sum_{\lambda\in\Lambda_n} \mathbf{D}_{AB}^{\lambda(n)} \leq \mathbf{I}_{AB}^{\otimes n},$$

and $\mathbf{I}_{AB}$ denoting the identity operator acting in $\mathbb{H}_{AB}$. Defining

$$\mathbf{D}_{AB}^{0(n)} := \mathbf{I}_{AB}^{\otimes n} - \sum_{\lambda\in\Lambda_n} \mathbf{D}_{AB}^{\lambda(n)}$$

yields a resolution of identity in $H_{AB}^{\otimes n}$. Hence, $\{\mathbf{D}_{AB}^{\lambda(n)}\}_{\lambda\in\Lambda_n\cup\{0\}}$ defines a POVM. An output $\beta \in \Lambda_n$ of a measurement described by this POVM, would lead Bob to conclude that the codeword was $\rho_{AB}^{\beta(n)}$, whereas the output 0 is interpreted as a failure of any inference. The encoding and decoding operations, employed to achieve reliable transmission of information by means of this protocol, together define a quantum code $\mathfrak{C}^{(n)}$ (of length n and size N_n), which is given by the triple $\mathfrak{C}^{(n)} := (\Lambda_n, \{\mathbf{E}_A^{\lambda(n)}\}, \{\mathbf{D}_B^{\lambda(n)}\})$, with N_n denoting its size, and $\{\mathbf{E}_A^{\lambda(n)}\}$ and $\{\mathbf{D}_B^{\lambda(n)}\}$, being the encoding and decoding maps employed.

Assuming equidistribution of messages among $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_{M_n}\}$, the average probability of error for the code $\mathfrak{C}^{(n)}$ is given by

$$\bar{\mathbb{P}}_{\text{err}}(\mathfrak{C}^{(n)}) \equiv \frac{1}{|\Lambda_n|} \sum_{\lambda\in\Lambda_n} (1 - \operatorname{tr}[(\Phi^{(n)} \otimes \mathfrak{I}_B^{\otimes n})(\rho_{AB}^{(n)})]) \tag{16.70}$$

If for a given $R > 0$, there exists a sequence $(\Lambda_n)_{n=1}^{+\infty}$ with

$$R \leq \liminf_{n\to+\infty} \frac{1}{n} \log(|\Lambda_n|),$$

and a sequence of codes $(\mathfrak{C}^{(n)})_{n=1}^{+\infty}$ of size $|\Lambda_n|$ such that

$$\lim_{n\to+\infty} \bar{\mathbb{P}}_{\text{err}}(\mathfrak{C}^{(n)}) = 0,$$

then R is said to be an achievable rate. We define the one-shot entanglement-assisted classical capacity of the long-term memory channel defined by (16.1) as

$$C_{\text{ea}}^{\text{lm},(1)}(\Phi) := \sup_{\omega_{AB}}\{\sup\{R \mid R \text{ is achievable}\}\}, \tag{16.71}$$

where the internal supremum is over the rates achievable under the choice of the initial shared state ω_{AB}. More generally, Alice and Bob may share indefinitely many copies of a pure state $\omega_{AB}^{\lambda,(m)}$ in $\mathbb{H}_{AB}^{\otimes m}$ for some given $m > 1$. In this case, Alice can perform a similar construction using encoding CPTP maps, $\mathfrak{E}_A^{(m)}$, which act in $\mathbb{H}_A^{\otimes m}$. In other words, she uses m-block encoding, and encodes a message $\lambda \in \Lambda_n$ by the state

$$(\rho_{AB}^{\lambda,m})^{(n)} := \rho_{\lambda,1}^{(m)} \otimes \cdots \otimes \rho_{\lambda,n}^{(m)} \in \mathcal{S}(\mathbb{H}_{AB}^{\otimes mn}),$$

where

$$\rho_{\lambda,k}^{(m)} = (\mathfrak{E}_{j_k}^{(m)} \otimes \mathfrak{I}_B^{\otimes m})\omega_{AB}^{m},$$

for $k = 1, 2, \ldots, n$ and $j_k \in \{1, 2, \ldots, J\}$. As before, Bob uses decoding POVM elements $\mathbf{D}_{AB}^{(n)}$, which are positive operators acting in $\mathbb{H}_{AB}^{\otimes mn}$, with $\sum_{i=1}^{|\Lambda_n|} \mathbf{D}_{AB}^{(n)} \leq \mathbf{I}^{\times mn}$.

The average probability of error for the resulting code $\mathfrak{C}^{(n)}$ is given by

$$\begin{aligned}\bar{\mathbb{P}}_{\mathrm{err},m}^{(m)} &\equiv \bar{\mathbb{P}}_{\mathrm{err}}(\mathfrak{C}^{(n)}) \\ &\equiv \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} (1 - \mathrm{tr}[\mathbf{D}_{\lambda,m}^{A,B;n}(\Phi^{(mn)} \otimes \mathfrak{I}_B^{\otimes mn})(\rho_{\lambda,m}^{AB;n})]).\end{aligned} \tag{16.72}$$

This gives rise to the m-shot entanglement-assisted classical capacity of the long-term memory channel defined by (16.1):

$$C_{\mathrm{ea}}^{\mathrm{lm},(m)}(\Phi) := \sup_{\omega_{AB}^{(m)}} \sup\{R \mid R \text{ is achievable}\}, \tag{16.73}$$

where the internal supremum is over the rates achievable under the choice of the initial shared state $\omega_{AB}^{(m)}$. Finally, the full entanglement-assisted classical capacity of Φ is given by

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) := \limsup_{m \to +\infty} \frac{1}{m} C_{\mathrm{ea}}^{\mathrm{lm},(m)}(\Phi). \tag{16.74}$$

16.3.1 Unconstrained case

Theorem 16.3.1. *Let $\mathbb{H}_A$ and $\mathbb{H}_B$ be two separable complex Hilbert spaces. The entanglement assisted classical capacity of a channel Φ with long-term memory, defined through* (16.74)*, is given by*

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) = \max_{\rho \in \mathcal{S}(\mathbb{H}_A)} \{\min\{I_m(\rho; \Phi_i), i = 1, 2, \ldots, M\}\} \tag{16.75}$$

with $I_m(\rho; \Phi_i) := H(\rho) + H(\Phi_i(\rho)) - H(\rho; \Phi_i)$, where $H(\rho; \Phi_i)$ denotes the entropy exchange and is defined as follows:

$$H(\rho; \Phi_i) := H((\Phi_i \otimes \mathfrak{R})\psi^{AR}), \tag{16.76}$$

with ψ^{AR} being a purification of ρ on a reference system R.

Theorem 16.3.1 will be proved in the following two results.

(A) Converse of Theorem 16.3.1

In the following, we prove that for any rate $R > C_{\text{ea}}^{\text{lm}}(\Phi)$, with $C_{\text{ea}}^{\text{lm}}(\Phi)$ given by (16.75), reliable entanglement-assisted transmission of classical information from Alice to Bob via the quantum channel Φ is impossible, regardless of the encoding used.

Suppose Alice and Bob share multiple copies of an entangled bipartite pure state $\omega_{AB}^{(m)} \in \mathcal{S}(\mathbb{H}_{AB}^{\otimes m})$, where m is a given positive integer. Then, given $n \in \mathbb{Z}$, Alice encodes her classical messages by applying chosen m-block encoding CPTP maps, n times, to her part of the shared state $(\omega_{AB}^{(m)})^{\otimes n}$. Here, we show that the average error probability of the corresponding code does not tend to zero as $n \to +\infty$, for any m and any choice of encoding maps. For notational simplicity, we will omit the label m and the superscript AB in the rest of this section.

Let $\sigma_\lambda^{(n)}(i) := \sigma_\lambda^1(i) \otimes \cdots \otimes \sigma_\lambda^n(i)$ denote Bob's final state, if the codeword

$$\rho_\lambda^{(n)} = \rho_\lambda^1 \otimes \cdots \otimes \rho_\lambda^n \in \mathfrak{B}(\mathbb{H}_{AB}^{\otimes mn}), \tag{16.77}$$

(16.77) corresponding to the message λ is transmitted through the ith branch, Φ_i, of the channel Φ. Here, $\sigma_\lambda^k(i) = (\Phi_i \otimes \mathfrak{I}_B)\rho_\lambda^k$, for $k = 1, 2, \ldots, n$,

$$\bar{\sigma}_\lambda^{(n)} = \sum_{i=1}^{M} p_i \sigma_\lambda^{(n)}(i); \quad \bar{\sigma}^{(n)}(i) = \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \sigma_\lambda^{(n)}(i)$$

and

$$\bar{\sigma}^k(i) = \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \sigma_\lambda^k(i), \quad \forall k = 1, 2, \ldots, n.$$

Then the average probability of error (16.72) equals

$$\bar{\mathbb{P}}_{\text{err}}^{(n)} := 1 - \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \text{tr}[\sigma_\lambda^{(n)} \mathbf{D}_\lambda^{(n)}]. \tag{16.78}$$

We also define the average probability of error corresponding to the ith branch of the channel as

$$\bar{\mathbb{P}}_{i,\text{err}}^{(n)} := 1 - \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \text{tr}[\sigma_\lambda^{(n)}(i) \mathbf{D}_\lambda^{(n)}] \quad \text{so that } \bar{\mathbb{P}}_{\text{err}}^{(n)} = \sum_{i=1}^{M} p_i \bar{\mathbb{P}}_{i,\text{err}}^{(n)}. \tag{16.79}$$

Let $X^{(n)}$ be a random variable with a uniform distribution over the set Λ_n, characterizing the classical message sent by Alice to Bob. Let $Y^{(n)}$ be the random variable corresponding to Bob's inference of Alice's message, when the codeword is transmitted through the ith branch of the channel. It is defined by the conditional probabilities

$$\Pr[Y^{(n)} = \beta | X^{(n)} = \alpha] = \text{tr}[\mathbf{F}_\beta^n(\Phi_i^{\otimes n} \otimes \mathfrak{I}_B^{\otimes n})(\rho_\alpha^{(n)})].$$

By Fano's inequality,

$$
\begin{aligned}
&h(\bar{\mathbb{P}}^{(n)}_{i,\mathrm{err}}) + \bar{\mathbb{P}}^{(n)}_{i,\mathrm{err}} \log(|\Lambda_n| - 1) \\
&\quad \geq H(X^{(n)}|Y_i^{(n)}) = H(X^{(n)}) - H(X^{(n)} \| Y_i^{(n)}).
\end{aligned} \tag{16.80}
$$

Here, $h(p) := -p\log p - (1-p)\log(1-p)$ denotes the binary entropy, $H(A) := -\sum_a p_a \log p_a$ denotes the Shannon entropy of a random variable A with probability mass function p_a, and $H(A|B)$, $H(A\|B)$ denote, respectively, the conditional entropy and the mutual information of two random variables A and B. Using the Holevo bound and the subadditivity of the von Neumann entropy, we have

$$
\begin{aligned}
&H(X^{(n)} \| Y_i^{(n)}) \\
&\quad \leq H\bigg(\frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} \sigma_\lambda^{(n)}(i)\bigg) - \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} H(\sigma_\lambda^{(n)}(i)) \\
&\quad \leq \sum_{k=1}^{n} H(\bar{\sigma}_k(i)) - \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} H(\sigma_{\lambda,k}(i)) \\
&\quad = \sum_{k=1}^{n} \chi\bigg(\bigg\{\frac{1}{|\Lambda_n|}, \sigma_{\lambda,k}(i)\bigg\}_{\lambda \in \Lambda_n}\bigg) \\
&\quad = \sum_{k=1}^{n} \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} H(\sigma_{\lambda,k}^{(n)}(i) \| \bar{\sigma}_k(i)) := \sum_{k=1}^{n} \mathbf{V}_k.
\end{aligned} \tag{16.81}
$$

In the above, the symbol $H(\rho\|\omega)$ denotes the quantum relative entropy of states ρ and ω. The expression $\mathbf{V}_k$ can be rewritten using Donald's identity:

$$
\sum_{\lambda \in \Lambda_n} p_\lambda H(\omega_\lambda \| \rho) = \sum_{\lambda \in \Lambda_n} p_\lambda H(\omega_\lambda \| \bar{\omega}) + H(\bar{\omega} \| \rho),
$$

where $\bar{\omega} = \sum_{\lambda \in \Lambda_n} p_\lambda \omega_\lambda$. We apply this with ρ replaced by

$$
\bar{\sigma}(i) = \frac{1}{n|\Lambda_n|} \sum_{k=1}^{n} \sum_{\lambda \in \Lambda_n} \sigma_{\lambda,k}(i),
$$

ω_λ replaced by $\sigma_{\lambda,k}(i)$, p_λ by $\frac{1}{|\Lambda_n|}$, and consequently, $\bar{\omega}$ replaced by $\bar{\sigma}_{\lambda,k}(i)$. Hence,

$$
\begin{aligned}
&\frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} H(\sigma_{\lambda,k}(i) \| \bar{\sigma}_k(i)) \\
&\quad = \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} H(\sigma_{\lambda,k}(i) \| \bar{\sigma}_k(i)) - H(\bar{\sigma}_{\lambda,k}(i) \| \bar{\sigma}(i)) \\
&\quad \leq \frac{1}{|\Lambda_n|} \sum_{\lambda \in \Lambda_n} H(\sigma_{\lambda,k}(i) \| \bar{\sigma}(i)),
\end{aligned}
$$

where we have used the nonnegativity of the quantum relative entropy. Inserting this into (16.81), we now have

$$\begin{aligned}\frac{1}{n}H(X^{(n)}\|Y_i^{(n)}) &\leq \frac{1}{n|\Lambda_n|}\sum_{k=1}^{n}\sum_{\lambda\in\Lambda_n} H(\sigma_{\lambda,k}(i)\|\bar{\sigma}(i))\\ &= \chi\left(\left\{\frac{1}{n|\Lambda_n|},\sigma_{\lambda,k}(i)\right\}_{\lambda,k}\right).\end{aligned}$$

The inequality (16.80) now yields

$$\begin{aligned}&h(\bar{\mathbb{P}}_{i,\mathrm{err}}^{(n)}) + \bar{\mathbb{P}}_{i,\mathrm{err}}^{(n)}\log(|\Lambda_n|)\\ &\quad\geq \log(|\Lambda_n|) - n\chi\left(\left\{\frac{1}{n|\Lambda_n|},\sigma_{\lambda,k}(i)\right\}_{\lambda,k}\right)\\ &\quad\geq \log(|\Lambda_n|) - nI_m(\rho;\Phi_i),\end{aligned}$$

where

$$\rho := \sum_{\lambda,k} p_{\lambda,k}\rho_{\lambda,k}^A$$

with $p_{\lambda,k} := \frac{1}{n|\Lambda_n|}$ for each λ and k, and $\rho_{\lambda,k}^A = \mathrm{tr}_B[\rho_{\lambda,k}]$, $k = 1,2,\ldots,n$.

However, since

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) \geq \min\{I_m(\rho;\Phi_i), i = 1,2,\ldots,M\}$$

and $R = \frac{1}{n}\log(|\Lambda_n|) > C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi)$, there must be at least one branch i such that

$$\bar{\mathbb{P}}_{i,\mathrm{err}}^{(n)} \geq 1 - \frac{C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) + 1/n}{R}. \tag{16.82}$$

We conclude from (16.79) and (16.82) that

$$\bar{\mathbb{P}}_{\mathrm{err}}^{(n)} \geq \left(1 - \frac{C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) + 1/n}{R}\right)\min\{p_i, i = 1,2,\ldots,M\}.$$

Hence, $\bar{\mathbb{P}}_{i,\mathrm{err}}^{(n)}$ does not tend to zero as $n \to +\infty$, which in turn implies that

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) \leq \max_{\rho}\{\min\{I_m(\rho;\Phi_i), i = 1,2,\ldots,M\}\}.$$

This proves the converse of Theorem 16.3.1. □

(B) Direct Part of Theorem 16.3.1
In the following, we prove that $C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi)$, defined by

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) = \limsup_{m\to+\infty} \frac{1}{m} C_{\mathrm{ea}}^{\mathrm{lm}(m)}(\Phi) \tag{16.83}$$

satisfies the lower bound

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) \geq \max_{\rho}\{\min\{I_m(\rho;\Phi_i), i = 1,2,\dots,M\}\}, \tag{16.84}$$

where the maximum is taken over all states $\rho \in \mathcal{S}(\mathbb{H}_A)$. To prove this, we employ the following result, which we proved in Theorem 16.2.5, for each of memoryless channel Φ_i for $i, 2, \dots, M$, which is restated below.

The product state capacity $C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)$ of a channel Φ with long-term memory is given by

$$C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi) = \sup_{\{q_j,\rho_j\}}\{\min\{\chi((q_j, \Phi_i(\rho_j)), i = 1,2,\dots,M\}\}, \tag{16.85}$$

where the supremum is taken over all finite ensembles of states $\rho_j \in \mathcal{S}(\mathbb{H}_A)$, chosen with probabilities q_j.

From the definition of the one-shot entanglement assisted capacity $C_{\mathrm{ea}}^{\mathrm{lm}(1)}(\Phi)$ defined in (16.85), it follows that

$$C_{\mathrm{ea}}^{\mathrm{lm}(1)}(\Phi) = \sup_{\pi_i,\mathfrak{E}_j,\omega_{AB}} \{\min\{(\chi(\{(\pi_j, (\Phi_i \otimes \mathfrak{I}_B)\rho_j^{AB})\}), i = 1,2,\dots,M\}\},$$

where (i) ω_{AB} is the bipartite entangled pure state, indefinitely many copies of which are shared by Alice and Bob and (ii) $\mathfrak{E}_j$ are encoding maps acting on $\mathcal{S}(\mathbb{H}_A)$, as described in earlier, i. e., $\rho_j^{AB} = (\mathfrak{E}_j \otimes \mathfrak{I}_B)\omega_{AB}$. Moreover, from the definition (16.73) of the m-shot entanglement assisted capacity it follows that

$$\begin{aligned} &C_{\mathrm{ea}}^{\mathrm{lm}(m)}(\Phi) \\ &\quad = \sup_{\pi_j^{m)},\mathfrak{E}_j,\omega_{AB}^{(m)}} \{\min\{\chi(\{\pi_j^{(m)}, (\Phi_i^{(m)} \otimes \mathfrak{I}_B)\rho_j^{AB,m}\}), i = 1,2,\dots,M\}\}. \end{aligned} \tag{16.86}$$

Now, consider a specific encoding ensemble $\{\pi_j^{(m)}, \mathfrak{E}_j^{(m)}\}$, where $a, b = 1,2,\dots,q$, for some integer q, and

$$\pi_{a,b}^{(m)} = \frac{1}{q^2}; \mathfrak{E}_{a,b}^{(m)} = \mathfrak{W}_{a,b}^{(m)}.$$

Here, $\mathfrak{W}_{a,b}^{(m)}$ denotes the discrete Weyl–Segal operators for a q-dimensional subspace $\mathbb{Q}_m$ of $\mathbb{H}_A^{\otimes m}$. Further, consider the codewords to be given by

$$\rho_{a,b}^{A,B,m} = (\mathfrak{W}_{a,b}^{(m)} \otimes \mathfrak{I}_B^{\otimes m})|\psi_m^{AB}\rangle\langle\psi_m^{AB}|$$

where $|\psi_m^{AB}\rangle$ denotes a maximally entangled state of rank q:

$$|\psi_m^{AB}\rangle := \frac{1}{\sqrt{q}}\sum_{k=1}^{q}|e_k^{(m)}\rangle \otimes |e_k^{(m)}\rangle,$$

where $\{|e_k^{(m)}\rangle\}_{k=1}^{q}$ is an orthonormal system of vectors in $\mathbb{Q}_m$. Hence,

$$C_{\mathrm{ea}}^{\mathrm{lm}(m)}(\Phi) \geq \min\left\{\chi\left(\left\{\frac{1}{q^2}, (\Phi_i^{\otimes m} \otimes \mathfrak{I}_B^{\otimes m})\rho_{a,b}^{AB,m}\right\}\right), i = 1, 2, \ldots, M\right\} \tag{16.87}$$

It follows that

$$\chi\left(\left\{\frac{1}{q^2}, (\Phi_i^{\otimes m} \otimes \mathfrak{I}_B^{\otimes m})\rho_{a,b}^{AB,m}\right\}\right) = I_m\left(\frac{\mathbf{P}^{(m)}}{\mathrm{tr}[\mathbf{P}^{(m)}]}; \Phi_i^{\otimes m}\right) \tag{16.88}$$

where $\mathbf{P}^{(m)}$ is the orthoprojection onto $\mathbb{Q}_m$. Further, it can be proved in that if $\mathbb{Q}_m$ is chosen to be the strongly δ-typical subspace for an arbitrary state $\rho^{\otimes m} \in \mathcal{S}(\mathbb{H}^{\otimes m})$, and $\mathbf{P}^{(m),\delta}$ is its orthoprojection, then

$$\lim_{\delta \to 0}\lim_{m \to +\infty}\frac{1}{m}I_m\left(\frac{\mathbf{P}^{m,\delta}}{\mathrm{tr}[\mathbf{P}^{m,\delta}]}; \Phi_i^{\otimes m}\right) = I_m(\rho; \Phi_i) \tag{16.89}$$

From (16.87), (16.88), (16.89) and (16.86) of the full entanglement-assisted capacity, it follows that

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) \geq \min\{I_m(\rho; \Phi_i), i = 1, 2, \ldots, M\}. \tag{16.90}$$

16.3.2 Energy constrained case

In this subsection, we assume that all Hilbert spaces are infinite-dimensional and investigate classical capacity $C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E)$ for the constrained channel Φ with long-term memory described by (16.1), where $\mathbf{H}$ is an $\mathfrak{H}$-operator on $\mathbb{H}_A$. This constraint is determined by the linear inequality

$$\mathrm{tr}[\rho\mathbf{H}] \leq E, \quad E > 0. \tag{16.91}$$

As mentioned earlier, we again denote the compact subset $\mathcal{K}_{\mathbf{H}}(E)$ of $\mathcal{S}(\mathbb{H}_A)$ (see (3.3)) as

$$\mathcal{K}_{\mathbf{H}}(E) = \{\rho \in \mathcal{S}(\mathbb{H}_A) \mid \mathrm{tr}[\rho\mathbf{H}] \leq E\}.$$

For the $\mathfrak{H}$-operator $\mathbf{H}$ on infinite-dimensional space $\mathbb{H}_A$ and any state $\rho \in \mathcal{S}(\mathbb{H}_A)$, the energy $\mathrm{tr}[\rho\mathbf{H}]$ (finite or infinite) is defined as $\sup_n \mathrm{tr}[\rho\mathbf{P}_n\mathbf{H}\mathbf{P}_n]$, where $\mathbf{P}_n$ is the finite-dimensional spectral projector of $\mathbf{H}$ corresponding to the interval $[0, n]$.

We impose the following linear constraint onto the input states $\omega^{(n)}$ of the channel $\Phi^{(n)}$,

$$\mathrm{tr}[\omega^{(n)}\mathbf{H}^{(n)}] \leq nE, \tag{16.92}$$

where

$$\mathbf{H}^{(n)} = \mathbf{H} \otimes \mathbf{I}_A \otimes \cdots \otimes \mathbf{I}_A + \cdots + \mathbf{I}_A \otimes \cdots \otimes \mathbf{I}_A \otimes \mathbf{H}. \tag{16.93}$$

Constraint (16.92) is equivalent to a similar constraint on input states of the channel $\Phi^{\otimes n} \otimes \mathfrak{I}_B^{\otimes n}$ with the constraint operator $\mathbf{H}_{AB}^{(n)} = \mathbf{H}^{(n)} \otimes \mathbf{I}_B^{(n)}$ on the composite Hilbert space $\mathbb{H}_{AB}^{\otimes n}$, where $\mathbf{I}_B^{(n)}$ is the identity operator on $\mathbb{H}_B^{\otimes n}$. Denote by $\mathcal{P}_{AB}^{(n)}$ the collection of ensembles $\pi^{(n)} = \{p_\lambda^{(n)}, \omega_\lambda^{(n)}\}$, where $\omega_\lambda^{(n)}$ are states of the form (13.73) satisfying

$$\sum_{\lambda \in \Lambda} p_\lambda^{(n)} \,\mathrm{tr}[\omega_\lambda^{(n)}\mathbf{H}_{AB}^{(n)}] \leq nE.$$

The classical capacity of the of the long-term memory channel Φ defined by (16.1) and under constraint (13.73), denoted by $C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E)$, is given in the following result.

Theorem 16.3.2. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be a quantum channel, and let $\mathbf{H}$ be an $\mathfrak{H}$-operator on $\mathbb{H}_A$. The entanglement-assisted classical capacity (finite or infinite) of the channel Φ with long-term memory described by* (16.1) *and with constraint* (13.73) *is given by the expression*

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E) = \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} \left(\min\{I_m(\rho; \Phi_i), i = 1, 2, \ldots, M\}\right) \tag{16.94}$$

where $I_m(\rho; \Phi_i)$ is the mutual information of Φ_i at the input state as defined in Definition 11.1.1.

We prove the above theorem via the following three lemmas. The proofs of Lemmas 16.3.3 and 16.3.4 are obtained based on necessary modifications of results found in Holevo and Shirokov [83].

Lemma 16.3.3. *Assume that* $\dim(\mathbb{H}_{A'}) < +\infty$. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be a channel with long-term memory described by* (16.1)*, and let $\mathbf{H}$ be an $\mathfrak{H}$-operator defined on $\mathbb{H}_A$. The entanglement-assisted classical capacity (finite or infinite) of the channel Φ with constraint* (13.73) *is given by the expression*

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E) \geq \sup_{\rho \in \mathcal{K}_{\mathbf{H}}(E)} \left(\min\{I_m(\rho; \Phi_i), i = 1, 2, \ldots, M\}\right). \tag{16.95}$$

Proof. Assuming that $\dim(\mathbb{H}_{A'}) < +\infty$, we follow the proof in Holevo and Shrikov [83] and observe the following:

1. Finite-dimensionality of the system $\mathbb{H}_{A'}$ implies finiteness of the output entropy of the channel Φ on the whole space of input states $\mathcal{S}(\mathbb{H}_A)$. That is, $H(\Phi(\rho)) < +\infty$ for all $\rho \in \mathcal{S}(\mathbb{H}_A)$;
2. Finiteness of $\mathrm{tr}[\rho\mathbf{H}]$ implies that all the eigenvectors of the state ρ belong to the domain of the operator $\sqrt{\mathbf{H}}$.
3. Finite-dimensionality of the system $\mathbb{H}_{A'}$ shows that for any finite-rank state ρ the restriction of the channel $\Phi^{\otimes n}$ to the support of the state $\rho^{\otimes n}$ acts as a finite-dimensional channel for each n;
4. If there are no states ρ satisfying the inequality $\mathrm{tr}[\rho\mathbf{H}] < E$ but there exists an infinite-rank state ρ_0 such that $\mathrm{tr}[\rho_0\mathbf{H}] = E$, then there is a sequence $(\rho_n)_{n=1}^{+\infty}$ of finite-rank states converging to ρ_0 such that $\mathrm{tr}[\rho_n\mathbf{H}] = E$ for which

   $$\liminf_{n\to+\infty} I_m(\rho_n, \Phi) \geq I_m(\rho_0, \Phi)$$

   by lower semicontinuity of the quantum mutual information.

With the observations above the rest of the proof of this lemma, follows similar to the direct part of that of Theorem 16.3.1. This proves the lemma. □

The following result removes the condition that $\dim(\mathbb{H}_{A'}) < +\infty$ from the above lemma.

Lemma 16.3.4. *Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be a quantum channel, and let $\mathbf{H}$ be an $\mathfrak{H}$-operator on $\mathbb{H}_A$. The entanglement-assisted classical capacity (finite or infinite) of the channel Φ with constraint* (13.73) *is given by the expression*

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E) \geq \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} (\min\{I_m(\rho; \Phi_i), i = 1, 2, \ldots, M\}). \tag{16.96}$$

Proof. We follow the proof provided in Holevo and Shirkov [83] below. Let $\Phi : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be an arbitrary channel. We prove that

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E) \geq \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} (\min\{I_m(\rho; \Phi_i), i = 1, 2, \ldots, M\})$$

as follows.

Let $(\mathbf{P}_n)_{n=1}^{+\infty}$ be sequence of finite-dimensional projectors on $\mathbb{H}_{A'}$ strongly converging to the unit operator $\mathbf{I}_{A'}$ on $\mathbb{H}_{A'}$. The channel Φ is approximated in the strong convergence topology by the sequence $(\Pi_n \circ \Phi)_{n=1}^{+\infty}$ with finite-dimensional output, where $\Pi_n(\sigma) = \mathbf{P}_n\sigma\mathbf{P}_n + (\mathrm{tr}[\sigma(\mathbf{I}_{A'} - \mathbf{P}_n)])\tau$ for all $\sigma \in \mathcal{S}(\mathbb{H}_{A'})$ and τ is a given state in $\mathbb{H}_{A'}$. Since the inequality "≥" in (16.95) is proved for a channel with finite-dimensional output (see Lemma 16.3.3), the chain rule for the entanglement-assisted capacity implies

$$C_{\text{ea}}^{\text{lm}}(\Phi;\mathbf{H},E) \geq C_{\text{ea}}^{\text{lm}}(\Pi_n \circ \Phi;\mathbf{H},E) \geq \min\{I_m(\rho;\Pi_n \circ \Phi_i), i = 1,2,\dots,M\}.$$

Lower semicontinuity of the function $\Phi \mapsto I_m(\rho;\Phi)$ in the strong convergence topology and the chain rule for quantum mutual information (see Proposition 11.2.3) imply

$$\lim_{n\to+\infty} I_m(\rho;\Pi_n \circ \Phi) = I_m(\rho;\Phi) \leq +\infty, \quad \forall\rho.$$

Hence, the inequality "≥" in (13.92) for the channel Φ follows from the above inequality. This proves the lemma. □

Lemma 16.3.5. *For each $i = 1,2,\dots,M$, let $\Phi_i : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_{A'})$ be a memoryless channel, and let $\mathbf{H}$ be an $\mathfrak{H}$-operator on $\mathbb{H}_A$. The entanglement-assisted classical capacity (finite or infinite) of channel Φ with long-term memory described by* (16.1) *and with constraint* (13.73) *is given by the expression*

$$C_{\text{ea}}^{\text{lm}}(\Phi;\mathbf{H},E) \leq \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} (\min\{I_m(\rho;\Phi_i), i = 1,2,\dots,M\}). \tag{16.97}$$

Proof. We prove the inequality (16.97). For the ensemble $\{p_\lambda,\omega_\lambda\}$ of encoded states in $\mathcal{S}(\mathbb{H}_{AA'})$, let $(\omega_\lambda)_A := \operatorname{tr}_{A'}[\omega_\lambda] = \sigma \in \mathcal{S}(\mathbb{H}_A)$, where $\operatorname{tr}_{A'}[\cdots]$ denotes the partial trace of $[\cdots]$ taken over $\mathbb{H}_{A'}$. By Lemma 13.4.4 and by the proof of the converse of Theorem 16.3.1, we have the following inequality:

$$\chi_{\Phi^{(n)}\otimes\mathfrak{I}_B^{\otimes n}}(\{p_\lambda^{(n)},\omega_\lambda^{(n)}\}) \leq \min\left\{I_m\left(\sum_\lambda p_\lambda^{(n)}(\omega_\lambda^{(n)})_A, \Phi_i^{\otimes n}\right), i = 1,2,\dots,M\right\}.$$

From (13.79), we have

$$C_{\text{ea}}^{\text{lm}}(\Phi;\mathbf{H},E) \leq \lim_{n\to+\infty} \frac{1}{n} \sup_{\pi^{(n)}\in\mathcal{P}_{AB}^{(n)}} \min\left\{I_m\left(\sum_\lambda p_\lambda^{(n)}(\omega_\lambda^{(n)})_A, \Phi_i^{\otimes n}\right), i = 1,2,\dots,M\right\}$$

Now

$$\begin{aligned}
&\sup_{\pi^{(n)}\in\mathcal{P}_{AB}^{(n)}} \left(\min\left\{I_m\left(\sum_\lambda p_\lambda^{(n)}(\omega_\lambda^{(n)})_A, \Phi_i^{\otimes n}\right), i = 1,2,\dots,M\right\}\right) \\
&\quad \leq \sup_{\rho^{(n)}:\operatorname{tr}[\rho^{(n)}\mathbf{H}^{(n)}]\leq nE} (\min\{(\rho^{(n)},\Phi_i^{\otimes n}), i = 1,2,\dots,M\}) \\
&\quad \equiv \min\{\bar{I}_m^{(n)}(\Phi_i), i = 1,2,\dots,M\}.
\end{aligned}$$

We claim that for each $i = 1,2,\dots,M$ the sequence $(\bar{I}_m^{(n)}(\Phi_i))_{n=1}^{+\infty}$ is additive. To show that, it suffices to prove that

$$\begin{aligned}
&\min\{\bar{I}_m^{(n)}(\Phi_i), i = 1,2,\dots,M\} \\
&\quad \leq n\min\{\bar{I}_m^{(1)}(\Phi_i), i = 1,2\dots,M\}
\end{aligned} \tag{16.98}$$

By subadditivity of quantum mutual information,

$$\min\{I_m(\rho^{(n)},\Phi_i^{\otimes n}), i=1,2,\dots,M\}$$
$$\leq \sum_j \min\{I_m(\rho_j^{(n)},\Phi_i), i=1,2,\dots\}$$

where $\rho_j^{(n)}$ are partial states, and by concavity,

$$\sum_j \min\{I_m(\rho_j^{(n)},\Phi_i), i=1,2,\dots,M\}$$
$$\leq n\min\left\{I_m\left(\sum_j \rho_j^{(n)},\Phi_i\right), i=1,2,\dots,M\right\}.$$

The inequality $\mathrm{tr}[\rho^{(n)}\mathbf{H}^{(n)}] \leq nE$ is equivalent to $\mathrm{tr}[(\frac{1}{n}\sum_{j=1}^n \rho_j^{(n)})\mathbf{H}] \leq E$, hence (16.98) holds. Thus,

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi;\mathbf{H},E) \leq \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} (\min\{I_m(\rho;\Phi_i), i=1,2,\dots,M\}).$$

This proves the lemma. □

Lemmas 16.3.3, 16.3.4 and 16.3.5 together prove that

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi;\mathbf{H},E) = \sup_{\rho\in\mathcal{K}_{\mathbf{H}}(E)} I_m(\rho;\Phi).$$

Consequently, Theorem 16.3.2 follows. This proves the theorem. □

The following corollary for unconstrained channels follows immediately from Theorem 16.3.2 by the fact that $\lim_{E\to+\infty}\mathcal{K}_{\mathbf{H}}(E) = \mathcal{S}(\mathbb{H}_A)$ in the $\|\cdot\|_1$-norm.

Corollary 16.3.6. *For each $i=1,2,\dots,M$, let $\Phi_i : \mathcal{S}(\mathbb{H}_A) \to \mathcal{S}(\mathbb{H}_B)$ be a memoryless quantum channel. Then the entanglement-assisted classical capacity (finite or infinite) of the unconstrained channel Φ with the long-term memory described by* (16.1) *is given by the expression*

$$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi) = \sup_{\rho\in\mathcal{S}(\mathbb{H}_A)} (\min\{I_m(\rho;\Phi_i), i=1,2,\dots,M\}). \tag{16.99}$$

Bibliography

[1] Aharonov, D., Kitaev, A., Nisan, N.: Quantum circuits with mixed states. In: *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computation (STOC)*, pp. 20–30 (1997).

[2] Arveson, W.B.: *A Short Course on Spectral Theory*. Springer, Berlin (2002).

[3] Attal, S.: Lecture 6. Quantum Channels (2015). http://math.univ-lyon1.fr/~attal/Quantum_Channels.pdf.

[4] Banach, S., Steinhaus, H.: Sur le principe de la condensation des singularities. *Fundam. Math.* **9**, pp. 50–61 (1927).

[5] Barnum, H., Nielsen, M.A., Schumacher, B.W.: Information transmission through a noisy quantum channel. *Phys. Rev. A* **57**(6), pp. 4153–4175 (1998).

[6] Banerjee, S., Roy, A.: *Linear Algebra and Matrix Analysis for Statistics*, 1st edn. Texts in Statistical Science. Chapman and Hall/CRC (2014). ISBN 978-1420095388.

[7] Bennett, C.H., DiVincenzo, D.P., Smolin, J.A., Wootters, W.K.: Mixed state entanglement and quantum error correction. *Phys. Rev. A* **54**, pp. 3824–3851 (1996).

[8] Bennett, C.H., Shor, P.W., Smolin, J.A., Thapliyal, A.V.: Entanglement- assisted classical capacity of noisy quantum channel. *Phys. Rev. Lett.* **83**, pp. 3081–3084 (1999).

[9] Bennett, C.H., Shor, P.W., Smolin, J.A., Thapliyal, A.V.: Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Trans. Inf. Theory* **48**(10), pp. 2637–2655 (2002).

[10] Billingsley, P.: *Ergodic Theory and Information*. John Wiley and Sons, New York–London–Sydney–Toronto (1965).

[11] Billingsley, P.: *Convergence of Probability Measures*. John Willey and Sons, New York–London–Sydney–Toronto (1968).

[12] Bjelakovic, I., Boche, H.: Ergodic Classical-quantum ahannels: structure and coding theorems. *IEEE Trans. Inf. Theory* **54**, p. 723 (2008).

[13] Bochner, S.: Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind. *Fundam. Math.* **20**, pp. 262–276 (1933).

[14] Bowen, G., Mancini, S.: Quantum channels with a finite memory (2004). arXiv:quant-ph/0305010v2.

[15] Bratteli, O., Robinson, D.W.: *Operator Algebras and Quantum Statistical Mechanics I*, 2nd edn. Springer (1987).

[16] Brocker, T., Werner, R.F.: *J. Math. Phys.* **36**, p. 62 (1995).

[17] Carathéodory, C.: Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen. *Math. Ann.* **64**(1), pp. 95–115 (1907) (in German).

[18] Caruso, F., Giovannetti, V., Lupo, C., Mancini, S.: Quantum channels and memory effects (2013). arXiv:1207.5435v4 [quant-ph].

[19] Cerf, N.J., Adami, C.: Negative Entropy and Information in Quantum Mechanics. *Phys. Rev. Lett.* **79**(26), pp. 5194–5197 (1997). arXiv:quant-ph/9512022.

[20] Cerf, N.J., Adami, C.: Quantum extension of conditional probability. *Phys. Rev. A* **60**(2), pp. 893–897 (1999). arXiv:quant-ph/9710001.

[21] Chang, M.-H.: *Stochastic Control of Hereditary Systems and Applications*. Series on Stochastic Modelling and Applied Probability, **59**. Springer, New York (2008).

[22] Chang, M.-H.: A survey on invariance and ergodicity of quantum Markov semigroups. *Stoch. Anal. Appl.* **32**(3), pp. 448–554 (2014).

[23] Chang, M.-H.: Recurrence and transience of quantum Markov semigroups. *Stoch. Anal. Appl.* **33**(1), pp. 123–198 (2015).

[24] Chang, M.H.: *Quantum Stochastics*. Cambridge University Press, London (2015).

[25] Clifton, R., Halvorson, H.: Bipartite-mixed-states of infinite-dimensional systems are generically nonseparable. *Phys. Rev. A* **61**, 012108 (2000).

https://doi.org/10.1515/9783110788105-017

[26] Clifton, R., Halvorson, H., Kent, A.: Non-local Correlations are Generic in Infinite-Dimensional Bipartite Systems. *Phys. Rev. A* **61**, 042101 (2000).

[27] Cohn, D.: *Measure Theory*. Springer (2013).

[28] Conway, J.B.: *A Course in Functional Analysis (GTM/96)*. Spring-Verlag, New York (1990).

[29] Cover, T., Thomas, J.: *Elements for Information Theory*. Wiley (1991).

[30] Davis, E.B.: *Quantum Theory of Open Systems*. Academic Press, London, New York (1976).

[31] Datta, N., Dorlas, T.: A Quantum Version of Feinstein's Lemma and its application to Channel Coding. In: *Proc. of Int. Symp. Inf. Theory*, Seattle, pp. 441–445 (2006).

[32] Datta, N., Dorlas, T.C.: Coding theorem for a class of quantum channels with long-term memory (2006). arXiv:quant-ph/0610049v2.

[33] Datta, N., Dorlas, T.: Classical capacity of quantum channels with general markovian correlated noise. *J. Stat. Phys.* **134**, p. 1173 (2009).

[34] Datta, N., Dorlas, T.C.: Entanglement assisted classical capacity of a class of quantum channels with long-term memory (2009). arXiv:0705.1465v2 [quant-ph].

[35] Datta, N., Suhov, Y., Dorlas, T.C.: The coding theorem for a class of quantum channels with long-term memory. Quantum Information Processing (2008).

[36] Del Palma, G., Trevisan, D., Giovannetti, V.: Gaussian states minimize output entropy of one-mode quantum Gaussian channels (2016). arXiv:1610.00970v1 [quant-ph].

[37] Devetak, I.: The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory* **51**, pp. 44–55 (2005).

[38] Dirac, P.A.M.: *The Principles of Quantum Mechanics*. International Series of Monographs on Physics. Oxford University Press, New York (1982).

[39] Dixmier, J.: *Von Neumann Algebras*. North Holland, Amsterdam (1981).

[40] Donald, M.J.: Further results on the relative entropy. *Math. Proc. Camb. Philos. Soc.* **101**, p. 363 (1987).

[41] Donald, M.J., Horodecki, M.: Continuity of relative entropy of entanglement. *Phys. Lett. A* **264**, p. 257 (1999).

[42] Dorlas, T., Morgan, C.: Classical capacity of quantum channels with memory (2009). arXiv:0902.2834v2 [quant-ph].

[43] Duan, L., Guo, G.: Reducing decoherence in quantum computer memory with all quantum bits coupling to the same environment (1998). arXiv:quant-ph/9612003v2.

[44] Einstein, A., Podosky, R.N.: Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, pp. 777–780 (1935).

[45] Eisert, J., Simon, C., Plenio, M.: On the quantification of entanglement in infinite dimensional system. journal of Physics A Mathematical and General **35**(17) (2002). 10.1088/0305-4470/35/17/307. arXiv:quant-ph/0112064v3.

[46] Erlijman, J., Farenick, D.R., Zeng, R.: Young's inequality in compact operators. *Oper. Theory, Adv. Appl.* **130**, pp. 171–184 (2001).

[47] Fan, K.: On a theorem of Weyl concerning the eigenvalue of linear transformation I. *Proc. Natl. Acad. Sci.* **35**(11), pp. 652–655 (1949).

[48] Fannes, A.: A continuity property of entropy density for spin lattice systems. *Commun. Math. Phys.* **31**, pp. 291–294 (1973).

[49] Fano, R.: *Transmission of information: a statistical theory of communications*. MIT Press, Cambridge, Mass. (1961). ISBN 978-0-262-56169-3. OCLC 804123877.

[50] Feinstein, A.: A new basic theorem of information theory. *IRE Trans. PGIT* **4**, pp. 2–22 (1954).

[51] Frechet, M.: Sur les ensembles de fonctions et les opérations linéaires. *C. R. Acad. Sci.* **144**, pp. 1414–1416 (1907) (in French).

[52] Ghne, O., Tuth, G.: Entanglement detection. *Phys. Rep.* **474**(1) (2009).

[53] Giovannetti, V.: A dynamical model for quantum memory channels. *J. Phys. A* **38**, 10989 (2005).

[54] Giovannetti, V., Guha, S., Lloyd, S., Maccone, L., Shapiro, J.H., Yuen, H.P.: Classical capacity of the lossy bosonic channel: the exact solution. *Phys. Rev. Lett.* **92**, 027902 (2004).

[55] Gallager, R.G.: *Information Theory and Reliable Communication*. Wiley, New Yourk (1968).

[56] Hall, B.C.: *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, 2nd edn. Graduate Texts in Mathematics, **222**. Springer (2015).

[57] Halmos, P.: *Measure Theory*. Graduate Text in Mathematics, **17**. Springer (1976).

[58] Halmos, P.: *A Hilbert Space Problems Workbook*, 2nd edn. Graduate Texts in Mathematics, **19**. Springer (1982).

[59] Halvorson, H., Clifton, R.: Generic Bell correlation between arbitrary local algebras in quantum field theory. *J. Math. Phys.* **41**, p. 1711 (2000).

[60] Hastings, M.B.: Superadditivity of communication capacity using entangled inputs. *Nat. Phys.* **5**, pp. 255–257 (2009).

[61] Hayashi, M.: *Quantum Information: An Introduction*. Springer, New York (2006).

[62] Hayashi, M., Nagaoka, H.: General formulas for capacity of classical-quantum channels. *IEEE Trans. Inf. Theory* **49**, pp. 1753–1768 (2003).

[63] Hayden, P., Jozsa, R., Petz, D., Winter, A.: Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Commun. Math. Phys.* **246**, pp. 359–374 (2004). arXiv:quant-ph/0304007.

[64] He, K., Hou, J., Li, M.: A von Neumann entropy condition for unitary equivalence of quantum states. *Appl. Math. Lett.* **25**(8), pp. 1153–1156 (2012).

[65] Helstrom, C.W.: *Quantum Detection and Estimation Theory*. Mathematics in Science and Engineering, **123**. Academic Press, London (1976).

[66] Hiai, F., Mosonyi, M., Petz, D., Byoeny, C.: Quantum f-divergences and error correction (2011). arXiv:1008.2529v5.

[67] Holevo, A.S.: Some estimates for information quantity transmitted by quantum communication channel. *Probl. Inf. Transm.* **9**, pp. 3–11 (1973).

[68] Holevo, A.S.: Bounds for the quantity of information transmitted by a quantum communication channel. *Probl. Inf. Transm.* **9**, pp. 177–183 (1973).

[69] Holevo, A.S.: The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory* **44**, pp. 269–273 (1998).

[70] Holevo, A.S.: *Statistical Structure of Quantum Theory*. Springer (2001).

[71] Holevo, A.S.: On entanglement-assisted classical capacity (2001). arXiv:quant-ph/0106075v2.

[72] Holevo, A.S.: Classical capacities of quantum channels with constrained inputs. *Probab. Theory Appl.* **48**(2), pp. 359–374 (2003).

[73] Holevo, A.S.: Entanglement-assisted capacity of constrained channels (2003). arXiv:quant-ph/0211170v2.

[74] Holevo, A.S.: On complementary channels and the additivity problem. *Probab. Theory Appl.* **51**, pp. 134–143 (2006).

[75] Holevo, A.S.: The entropy gain of infinite-dimensional quantum channels. *Dokl. Math.* **82**(2), pp. 730–731 (2010). arXiv:1003.5765.

[76] Holevo, A.S.: On the Choi-Jamiolkowski Correspondence in Infinite Dimensions (2011). arXiv:1004.0196v3 [math-ph].

[77] Holevo, A.S.: *Quantum Systems, Channels, Information. A Mathematical Introduction*, DeGruyter, Berlin (2012).

[78] Holevo, A.S.: On the constrained classical capacity of infinite dimensional covariant quantum channels (2015). arXiv:1409.8085v3 [quant-ph].

[79] Holevo, A.S., Shirokov, M.E.: On Shor's channel extension and constrained channels. *Commun. Math. Phys.* (2003). arXiv:quant-ph/0306196.

[80] Holevo, A.S., Shirokov, M.E.: Continuous ensembles and the χ-capacity of infinite dimensional channels. *Probab. Theory Appl.* **50**(1), pp. 98–114 (2005). arXiv:quant-ph/0408176.

[81] Holevo, A.S., Shirokov, M.E.: On approximation of quantum channels (2006). arXiv:quant-ph/0711.2245.

[82] Holevo, A.S., Shirokov, M.E.: Mutual and coherent information for infinite dimensional quantum channels. *Probl. Inf. Transm.* **46**(3), pp. 201–218 (2010).

[83] Holevo, A.S., Shirokov, M.E.: On classical capacities of infinite- dimensional quantum channels. *Probl. Inf. Transm.* **49**(1), pp. 15–31 (2013). arXiv:1210.6926.

[84] Holevo, A.S., Shirokov, M.E., Werner, R.F.: On the notion of entanglement in Hilbert space. *Russ. Math. Surv.* **60**(2), pp. 153–154 (2005). arXiv:quant-ph/0504204.

[85] Holevo, A., Shirokov, M.E., Werner: Separability and Entanglement Breaking in Infinite Dimensions (2005). quant-ph/0504204v1.

[86] Holevo, A.S., Werner, R.F.: Evaluating capacities of bosonic Gaussian channels. *Phys. Rev. A* **63**, 032312 (2001).

[87] Horodecki, P., Cirac, J.I., Lewenstein, M.: Bound Entanglement for Continuous Variables is a Rare Phenomenon. In: *Quantum Information with Continuous Variables* (2003).

[88] Horodecki, P., Lewenstein, M.: *Phys. Rev. Lett.* **85**, p. 2657 (2000).

[89] Horodecki, R., Horodecki, P., Horodecki, M., Horodecki, K.: Quantum entanglement. *Rev. Mod. Phys.* (2009).

[90] Hou, J., Qi, X.: Fidelity of states in infinite dimensional quantum systems (2011). arXiv:1107.0354v1 [quant-ph].

[91] Hu, S., Yu, Z.: The necessary and sufficient conditions for separability of bipartie pure states of infinite dimensional Hilbert spaces (2007). arXiv:0704.0969v1 [quant-ph].

[92] Hu, S., Yu, Z.: Note on Schmidt decomposition in infinite dimensional Hilbert spaces (2007). arXiv:0705.1694v1 [quant-ph].

[93] Jencova, A., Ruskai, M.B.: A unified treatment of convexity of relative entropy and related trace functions, with conditions for equality (2010). arXiv:0903.2895.

[94] Jencova, A.: Reversibility conditions for quantum operations (2012). arXiv:1107.0453.

[95] Jencova, A., Petz, D.: Sufficiency in quantum statistical inference. *Commun. Math. Phys.* **263**, pp. 259–276 (2006).

[96] Kadison, R.: A generalized Schwarz inequality and algebraic invariants for operator algebras. *Ann. Math.* **56**(2), pp. 494–503 (1952).

[97] King, C.: Remarks on additivity conjectures for quantum channels. *Contemp. Math.* **529**, p. 177 (2010).

[98] Kraus, K.: *States, Effects and Operations*. Springer, Berlin (1983).

[99] Kretschmann, D., Schlingemann, D., Werner, R.F.: A Continuity Theorem for Stinespring's Dilation (2007). arXiv:0710.2495.

[100] Kretschmann, D., Werner, R.F.: Quantum Channels with Memory. *Phys. Rev. A* **72** (2005).

[101] Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann. Math. Stat.* **22**(1), pp. 79–86 (1951).

[102] Kuznetsova, A.A.: Conditional entropy for infinite-dimensional quantum systems. *Theory Probab. Appl.* **55**, pp. 709–717 (2011).

[103] Lanford, O., Robinson, D.W.: Mean entropy of states in quantum statistical mechanics. *J. Math. Phys.* **9**, p. 1120 (1968).

[104] Lieb, E.H.: Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Adv. Math.* **11**, pp. 267–288 (1973).

[105] Lieb, E.H., Ruskai, M.B.: Proof of the strong subadditivity of quantum mechanical entropy. *J. Math. Phys.* **14**, p. 1938 (1973).

[106] Lindblad, G.: Entropy, Information and Quantum Measurements. *Commun. Math. Phys.* **33**, pp. 305–322 (1973).

[107] Lindblad, G.: Expectation and entropy inequlities for finite quantum systems. *Commun. Math. Phys.* **39**, pp. 111–119 (1974).

[108] Lindblad, G.: Completely positive maps and entropy inequalities. *Commun. Math. Phys.* **40**, pp. 147–151 (1975).
[109] Lloyd, S.: Capacity of the noisy quantum channel. *Phys. Rev. A* **55**(3), pp. 1613–1622 (1997).
[110] Luo, S., Li, N., Cao, X.: Relative entropy between quantum ensembles. *Period. Math. Hung.* (2009).
[111] Macchiavello, C., Palma, G.M.: Entanglement–enhanced information transmission over a quantum channel with correlated noise. *Phys. Rev. A* **65**, 050301 (2002).
[112] McMillan, B.: The basic theorems of information theory. *Ann. Math. Stat.* **24**, pp. 196–219 (1953).
[113] Majewski, A.W.: On entanglement of formation. *J. Phys. A* **35**(1), pp. 123–134 (2002).
[114] Moore, E.H.: On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.* **26**(9) (1920).
[115] Mosonyi, M., Ogawa, T.: Strong converse exponent for classical-quantum coding (2015). arXiv:1409.3562v5 [quant-ph].
[116] Nielsen, M.A., Chuang, I.L.: *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK (2000).
[117] Norris, J.R.: *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (1997).
[118] Ogawa, T., Nagaoka, H.: A new proof of the channel coding theorem via hypothesis testing in quantum information theory. In: *Proc. 2002 IEEE International Symposium on Information Theory*, p. 73 (2002).
[119] Ogawa, T., Sasaki, A., Iwamoto, M., Yamamoto, H.: Quantum secret Sharing Schemes and reversibility of auantum operations. *Phys. Rev. A* **72**, 032318 (2005).
[120] Oreshekov, O., Calsamiglia, J.: Distinguishablity between ensemble in quantum states. *Phys. Rev. A* (2009).
[121] Ohya, M., Petz, D.: *Quantum Entropy and Its Use*. Springer, Berlin (1993).
[122] Parthasarathy, K.: *Probability Measures on Metric Spaces*. Academic Press, New York and London (1967).
[123] Paulsen, V.: *Completely Bounded Maps and Operator Algebras*. Cambridge University Press (2003).
[124] Penrose, R.: A generalized inverse for matrices. *Proc. Camb. Philos. Soc.* **51**(3), pp. 406–413 (1955).
[125] Petz, D.: Sufficiency of channels over von Neumann algebras. *Q. J. Math. Oxford Ser. 2* **39**, pp. 97–108 (1988).
[126] Petz, D.: Monotonicity of quantum relative entropy revisite. *Rev. Math. Phys.* **15**, pp. 79–91 (2003).
[127] Protasov, V.Y., Shirokov, M.E.: Generalized compactness in linear spaces and its applications. *Sb. Math.* **200**(5), pp. 697–722 (2009). arXiv:1002.3610 [math-ph].
[128] Reed, M., Simon, B.: *Methods of Modern Mathematical Physics. Vol I. Functional Analysis*. Academic Press Inc. (1980).
[129] Riesz, F.: Sur une espèce de géométrie analytique des systèmes de fonctions sommables. *C. R. Acad. Sci.* **144**, pp. 1409–1411 (1907) (in French).
[130] Robinson, D.W., Ruelle, D.: *Commun. Math. Phys.* **5**, p. 288 (1967).
[131] Rockafellar, R.T.: *Convex Analysis*. Tyrrell (1970).
[132] Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*. Springer (1998).
[133] Rudin, W.: *Real and Complex Analysis*. McGrow-Haill, New York (1970).
[134] Rudin, W.: *Functional Analysis*, 2nd edn. Science/Engineering/Mathematics, McGraw-Hill, New York (1991).
[135] Ruskai, M.B.: Inequalities for quantum entropy: a review with conditions for equality. *J. Math. Phys.* **43**, pp. 4358–4375 (2002). arXiv:quant-ph/0205064.

[136] Rybár, T., Ziman, M.: Quantum finite-depth memory channel: case study. *Phys. Rev. A* **80**, 042306 (2009).
[137] Sakai, S.: *C^*-algebras and W^*-algebras*. Classics in Mathematics. Springer, Berlin (1998).
[138] Schumacher, B.: Quantum coding. *Phys. Rev. A* **51**(4), pp. 2738–2747 (1995).
[139] Schumacher, B., Westmoreland, M.D.: Optimal signal ensembles (1999). arXiv:quant-ph/9912122v1.
[140] Shannon, C.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, pp. 379–423 (1948).
[141] Shirokov, M.E.: Entropic characteristics of subset of states (2005). arXiv:quant-ph/0510073v2.
[142] Shirokov, M.E.: The Holevo capacity of infinite dimensional channels and the additivity problem (2005). arXiv:quant-ph/0408009.
[143] Shirokov, M.E.: The Convex Closure of the Output Entropy of In- finite Dimensional Channels and the Additivity Problem (2006). arXiv:quant-ph/0608090.
[144] Shirokov, M.E.: Continuity of the von Neumann Entropy. *Commun. Math. Phys.* **296**(3), pp. 625–654 (2010).
[145] Shirokov, M.E.: The output entropy of quantum channels and quantum operations (2010). arXiv:1002.0230.
[146] Shirokov, M.E.: The properties of the set of quantum states and their application to construction of entanglement monotones. *Izv. Math.* **74**(4), pp. 849–882 (2010).
[147] Shirokov, M.E.: Monotonicity of the Holevo quantity; a necessary condition for equality in terms of a channel and its applications (2011). arXiv:1106.3297.
[148] Shirokov, M.E.: Reversibility condition for quantum channels and their applications (2013). arXiv:1203.0262v3.
[149] Shirokov, M.E.: Monotonicity of the Holevo quantity: a necessary condition for equality in terms of a channel and its applications (2013). 1106.3297v6.
[150] Shirokov, M.E.: Estimates for discontinuity jumpsf information characteristics of quantum systems and channels. *Probl. Inf. Transm.* (2016).
[151] Shirokov, M.E.: Energy-constrained diamond norms and their use in quantum information theory (2017). arXiv:1706.0036v2 [quant-ph].
[152] Shirokov, M.E.: Strong convergence of quantum channels: continuity of the Stinespring dilation and discontinuity of unitary dilation (2018). arXiv:1712.03219v5 [quant-ph].
[153] Shirokov, M.E.: Strong* convergence of quantum channels (2018). arXiv:arKiv:1802.05632v2 [quant-ph].
[154] Shirokov, M.E.: Uniform continuity bounds for information characteristics of quantum channels depending on input dimension and on input energy (2018). arXiv:1610.08870v5 [quant-ph].
[155] Shirokov, M.E.: On extension of quantum channels and operations to the space of relatively bounded operators (2019). arXiv:1903:06086v1.
[156] Shirokov, M.E.: Strong convergence of quantum channel: continuity of the Stinespring dilation and discontinuity of the unitary dilation. *J. Math. Phys.* (2020).
[157] Shirokov, M.E., Bulinski, A.V.: On quantum channels and operations that preserving finiteness of the von Neumann entropy (2020). arXiv:2004.03586v1 [quan-ph].
[158] Shirokov, M.E., Holevo, A.S.: On approximation of infinite dimensional quantum channels. *Probl. Inf. Transm.* **44**(2), pp. 3–22 (2008). arXiv:quant-ph/0711.2245.
[159] Shor, P.W.: Additivity of the classical capacity of entanglement breaking quantum channel. *J. Math. Phys.* **43**, pp. 4334–4340 (2002).
[160] Shor, P.W.: Equivalence of additivity questions in quantum information theory. *Commun. Math. Phys.* **246**, pp. 453–472 (2004). LANL e-print: arXiv:quant-ph/0305035.
[161] Schrodinger, E.: Discussion of probability relations between separated systems. *Proc. Camb. Philos. Soc.* **31**, pp. 555–563 (1935).

[162] Schumacher, B., Nielsen, M.A.: Quantum data processing and error correction. *Phys. Rev. A* (1996).
[163] Schumacher, B., Westmoreland, M.D.: Sending classical informationvia noisy quantum channels. *Phys. Rev. A* **56**(1), pp. 131–138 (1997).
[164] Simon, B.: Convergence theorem for entropy. Appendix in Lieb E. H., Ruskai M. B.: Proof of the strong suadditivity of quantum mechanical entropy. *J. Math. Phys.* **14**, p. 1938 (1973).
[165] Simons, S.: *Minimax and Monotonicity*. Springer, Berlin (1998).
[166] Steinitz, E.: Bedingt konvergente Reihen und konvexe Systeme. *J. Reine Angew. Math.* **1913**(143), pp. 128–175 (1913).
[167] Stinespring, W.F.: Positive functions on C^*-algebras. *Proc. Am. Math. Soc.* **6**, pp. 211–216 (1955).
[168] Takesaki, M.: *Theory of Operator Algebras I*. Encyclopaedia of Mathematical Sciences, **124**. Springer (1979).
[169] Uhlmann, A.: The "transition probability" in the state space of a *-algebra. *Rep. Math. Phys.* **9**(2), pp. 273–279 (1976).
[170] Uhlmann, A.: *Open Syst. Inf. Dyn.* **5**, p. 209 (1998).
[171] van Neerven, J.: Functional Analysis. arXiv:2112.11162v2 [math.FA]. To be published by Cambridge University Press (2022).
[172] von Neumann, J.: *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, Princeton, N.J. (1996).
[173] Watrous, J.: *The Theory of Quantum Information*. Cambridge University Press, London (2018).
[174] Wehrl, A.: Three theorems about entropy and convergence of density matrices. *Rep. Math. Phys.* **10**, p. 159 (1976).
[175] Wehrl, A.: General properties of entropy. *Rev. Mod. Phys.* **50**, pp. 221–250 (1978).
[176] Werner, R.F.: Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A* **40**, p. 4277 (1989).
[177] Wheeden, R.L., Zygmund, A.: *Measure and Integral: An Introduction to Real Analysis*. Marcel Dekker, Inc., New York and Basel (1977).
[178] Wilde, M.: *Quantum Information Theory*. Cambridge University Press, London (2013).
[179] Wilde, M.: Multipartite quantum correlations and local recoverability. *Proc. R. Soc. A* **471**, p. 2177 (2015). arXiv:1412.0333.
[180] Winter, A.: *Coding Theorems of Quantum Information Theory*, Dissertation zur Erlangung des Doktorgrades, vorgelegt der Fakulty at fur Mathematik, Universitat Bielefeld, April (1999).
[181] Winter, A.: Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory* **45**, pp. 2481–2485 (1999).
[182] Yosida, K.: *Functional Analysis*, 5th edn. Springer, Berlin, Heidelberg, New York (1980).
[183] Zhang, L., Wu, J.: Von Neumann entropy-preserving quantum operations. *Phys. Lett. A* **375**(47), pp. 4163–4165 (2011). arXiv:1104.2992.
[184] Zorn, M.: A remark on method in transfinite algebra. *Bull. Am. Math. Soc.* **41**(10), pp. 667–670 (1935).

Index

$\mathbf{0}_{\mathbb{X}}$ 10
$]-\infty, a]$ 1
$[a, b]$ 1
$[a, b[$ 1
$]a, b]$ 1
$[b, +\infty[$ 1
$\langle \cdot, \cdot \rangle_2$ 5
$\langle \cdot, \cdot \rangle_{\mathbb{C}^n}$ 98
$\langle \cdot, \cdot \rangle_{\mathbb{H}}$ 2
$\langle \cdot, \cdot \rangle_{l^2}$ 5
$\langle \cdot, \cdot \rangle_{L^2}$ 5
$|z|$ 1
$\| \cdot \|_{\diamond}$ 149
$\| \cdot \|_{\mathbb{B}}$ 3
$\| \cdot \|_{\mathrm{cb}}$ 150
$\| \cdot \|_{HS}$ 29
$\| \cdot \|_{\infty}$ 10
$\| \cdot \|_{\mathbb{H}}$ 2
$\|\Phi\|_{\rho}$ 321
$\|\mathbf{T}\|_1$ 26
$\|\mathbf{T}\|_{\mathrm{HS}}$ 29
$\|\mathbf{T}\|_{\infty}$ 10
$\bar{A}$ 3
$\mathcal{A}_*$ 52
$\mathcal{A}'$ 39
$(\mathcal{A}, \rho)$ 60
$\mathbf{a}(E)$ 59
$\mathfrak{B}_+(\mathbb{H})$ 14
$\mathcal{B}(\mathbb{R})$ 61
$\mathfrak{B}(\mathbb{X})$ 10
$\mathfrak{B}(\mathbb{X}, \mathbb{Y})$ 10
$\mathbb{C}$ 1
$C_\chi(\Phi)$ 304
$C_\chi(\Phi; \mathcal{A})$ 304
$C_{\mathrm{ea}}(\Phi)$ 359, 361, 459
$C_{\mathrm{ea}}(\Phi; \mathbf{H}, E)$ 368
C^*-algebra
– unital 37
$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi)$ 461
$C_{\mathrm{ea}}^{\mathrm{lm}}(\Phi; \mathbf{H}, E)$ 467
$C_{\mathrm{prod}}^{\mathrm{lm}}(\Phi)$ 443
$\mathbb{C}^N$ 4
$\mathfrak{C}^{(n)}$ 357
$\mathbb{C}\phi$ 20
$C(K)$ 38
$\mathrm{co}(\mathcal{A})$ 81
$\mathrm{co}(\mathrm{epi}(f))$ 91
$\mathrm{co}(f)$ 91
$\overline{\mathrm{co}}(\mathcal{A})$ 81
$\dim(\mathbb{H})$ 7
$\mathrm{dom}(\mathbf{T})$ 7
$\mathrm{epi}(f)$ 91
$F_{\mathrm{ef}}(\rho, \Phi)$ 147
$\bar{F}(\{p_i, \rho_i\}, \Phi)$ 147
$\mathrm{graph}(\mathbf{T})$ 8
$\mathbb{H}$ 2
$\mathfrak{H}$-operator 178
$\mathbb{H}_A \otimes \mathbb{H}_B$ 62
$\mathfrak{H}\mathfrak{S}(\mathbb{H})$ 29
$I_c(\rho, \Phi)$ 289
$\mathbf{I}_{\mathbb{H}}$ 4
$I_m(p^{(n)}, \rho^{(n)}, \mathbf{D}^{(n)})$ 357
$\mathbf{I}_{\mathbb{X}}$ 10
$\mathfrak{I}(z)$ 1
ι 11
$\mathcal{K}_{\mathbf{H}}(h)$ 78
$(\ker(\Upsilon))^{\perp}$ 119
$\ker(\mathbf{T})$ 8
$\mathfrak{L}(\mathbb{X}, \mathbb{Y})$ 8
$\mathfrak{L}_c(\mathbb{H}, \mathbb{K})$ 22
$\mathfrak{L}_+(\mathbb{H})$ 14
$\mathfrak{L}_c(\mathbb{H})$ 22
$l_{\infty}(\mathbb{N}; \mathbb{C})$ 56
$l^2(\mathbb{N}; \mathbb{C})$ 5
$L^2(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu; \mathbb{C})$ 5
$\lambda_{\dagger}(\sigma)$ 179
$\lambda^{\dagger}(\mathbf{H})$ 179
$\mathcal{M}_n$ 98
$\mathbb{M} \oplus \mathbb{N}$ 17
$\mathcal{M}(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 58
$\mathbb{N}$ 2
$\mathcal{O}_{\mathbb{H}}(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ 59
$\Omega(\Phi, \mathcal{A})$ 321
$\mathbf{p}_{\phi}$ 20
$\mathcal{P}_{\rho}(\mathcal{S}(\mathbb{H}))$ 83
$\mathbf{P}_{\varphi}$ 54
$\mathbf{P}_a$ 60
$\mathcal{P}_{\mathcal{A}}(\mathcal{S}(\mathbb{H}))$ 83
$\mathcal{P}_{\mathrm{dis}}(\mathcal{S}(\mathbb{H}))$ 72
$\mathbb{P}_{\mathrm{err}}(\mathfrak{C}^{(n)})$ 357
$\mathbb{P}_{\mathrm{err}}(n, N)$ 357
$\mathcal{P}^{\mathrm{con}}(\mathcal{S}(\mathbb{H}))$ 73
$\mathcal{P}^{\mathrm{dis}}(\mathcal{S}(\mathbb{H}))$ 73
$P(\rho, \mathbf{a}, E)$ 60

https://doi.org/10.1515/9783110788105-018

$\mathrm{POVM}^1_{\mathbb{H}}(\mathbb{X})$ 58
$\mathrm{POVM}_{\mathbb{H}}(\mathbb{X})$ 58
$\mathbf{Q}_n$ 77, 178
$\mathfrak{QC}(A)$ 119
$\mathfrak{QC}(A,B)$ 119
$\mathfrak{QO}(A)$ 118
$\mathfrak{QO}(A,B)$ 118
$\mathbb{R}$ 1
range($\mathbf{T}$) 8
$S^\perp$ 18
$\mathbf{S} \circ \mathbf{T}$ 9
$\mathcal{S}(\mathfrak{A}^{(\infty)})$ 409
$\mathcal{S}(\mathfrak{A}^{(n)})$ 409
span(A) 4
span(S) 18
supp(μ) 72
supp($\mathbf{T}$) 8
$(\mathbf{T} - \lambda \mathbf{I})^{-1}$ 15
$\mathfrak{T}_+(\mathbb{H})$ 71
$\mathfrak{T}_{\leq 1}(\mathbb{H})$ 71
$\mathbf{T}^*$ 8
$\mathbf{T}^{-1}$ 8
$\mathfrak{T}(\mathbb{H})$ 26
tr[$\mathbf{T}$] 25
$(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ 5
$(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mathbf{a})$ 59
χ-capacity
– additive 330
– subadditive 330
– superadditivity 330
$\chi_\Phi(\mu)$ 296
$\Gamma(\mathcal{A})$ 177
$\hat{\Phi}$ 132
μ_v 59
$\Phi_{(\infty)}$ 409
$\Phi_C^{(n)}(\rho^{(n)}_{\Lambda_n,A})$ 422
$\Phi^{(n)}(\rho^{(n)}_{\Lambda_n,A})$ 409
π 38
$(\pi, \mathbb{H})$ 56
$(\pi, \mathbb{H}, v_\pi)$ 57
$(\pi_1, \mathbb{H}_1) \cong (\pi_2, \mathbb{H}_2)$ 56
$\Pi^{\pm}_{i,j}$ 444
$\Pi(\mathcal{A}, \mathfrak{B}(\mathbb{H}))$ 56
$[\pi(\mathcal{A})\mathbb{H}]$ 57
$\psi \perp \phi$ 4
$\psi(a)$ 60
$\mathbb{R}(z)$ 1
$\rho \perp \sigma$ 134
$\rho(\mathbf{T})$ 15
σ-algebra
– Borel 61
$\sigma_c(\mathbf{T})$ 15
$\sigma_p(\mathbf{T})$ 15
$\sigma_r(\mathbf{T})$ 15
$\sigma(\mathbf{T})$ 15
*-homomorphism 56
*-automorphism 39
*-isomorphism 39
$\vee_\alpha \mathbf{a}_\alpha$ 36

algebra
– *- 39
– C^*- 37
– quasi-local 393
– von Neumann 39
average state
– optimal 307

backaction 60
ball
– unit closed 24
barycenter 296
basis
– complete orthonormal 4
– orthonormal 4
block 342
block code 342

capacity
– entanglement-assisted classical 368, 459
– Holevo χ- 304
– product state 443
chain
– Markov 405
channel
– $\mathcal{A}$-constrained 303
– complementary 132
– complementary quantum 132
– concatenated 390
– extended quantum 120
– FF 329
– FI 329
– forgetful 397
– IF 330
– II 330
– isometric 120
– memory 383, 388

– memoryless 340
– quantum 119
– reversible 131
– reversible quantum 131
– reversing 131
– sufficient 131
– unitary 120
classical capacity
– constrained 357
– constrained entanglement assisted 467
– entanglement assisted 461
closure 3, 18
code 346, 356, 458
– quantum error correction 339
codimension 17
coefficient
– decrease 179
– increase 179
– Schmidt 244
coefficients
– Schmidt 244
coisometry 16
commutant 39
compact
– relatively 3
concavity 174
– generalized 174
condition
– consistency 388
– non-anticipatory 388
cone
– positive 71
conjugate
– complex 1
conjugation
– complex 2
continuity
– σ-weak 49
– absolute 203
– ultra-weak 41
convergence
– σ-strong 35
– σ-weak 35
– norm 35
– strong 6, 14, 35, 161
– strong* 35, 165
– uniform 161
– weak 6, 35
– weak* 35
convex hull
– σ 92
convexity
– generalized 208
correlation
– classical 263
covering
– open 3
Criterion
– Prohorov's Compactness 74

decomposable
– countably 263
decomposition
– canonical convex 55
– polar 30
– Schmidt 63, 244
– spectral 55, 130
dilation
– Stinespring 127
– unitary 125
direct sum 17
distance
– Bures 145, 151
– energy-constrained Bures 156
distribution
– stationary 406
divergence
– Kullback–Leibier 204
domain 7

effect 341
element
– normal 59
ensemble
– continuous 73
– discrete 73
– optimal discrete 305
– optimal generalized 306
– optimal signal 305
entanglement
– entropy 266
entanglement formation 272
entropy
– conditional 214
– relative 180, 210
– Shannon 170
– von Neumann 170, 196
entropy loss 201

EoE 266
EoF 272
equation
– Parsaval 5
equivalence
– isometric 119
– isometrical 119
– unitary 56
event 60
– quantum 60

face 232
fidelity 151
– channel 146
– ensemble average 147
– entanglement 147
fidelity of states 133
formula
– polarization 10
function
– χ- 295, 296
– μ-convex 89
– σ-convex 89
– bounded 88
– lower bounded 88
– semi-bounded 88
– unconstrained χ- 295
– upper bounded 88
functional
– bounded linear 42
– linear 42
– normal 49
– sublinear 42
– support 44

graph 8
group
– translation 394

Hilbert–Schmidt norm 29
Holevo quantity
– mean 429
homomorphism
– * 38
hull
– closed convex 81
– convex 81, 91
hyperline 44
hyperplane 44

identity
– Donald 313
– resolution of 61
inequality
– Cauchy-Schwarz 6
– discrete Jensen 90
– Fano 439, 454
– integral Jensen 89
– Kadison–Schwarz 114
– Klein 171
– Schwarz 114
inequality-Young's 27
information
– coherent 289, 324
– mutual 278
– Shannon 357
inner product
– Hilbert–Schmidt 29
integral
– Bochner 82
interval 2
invariance
– unitary 173
inverse 8
involution 37
isometry 16, 30
– partial 30, 119
– surjective 16
isomorphism 4

lemma
– Zorn's 35
linear span 18

manifold
– linear 4
map
– bijective 8
– bounded linear 10
– linear 8
Markov chain
– aperiodic 407
– ergodic 407
– irreducible 406
– period of 407
– periodic 407
– reducible 407
– stationary distribution 406
– time-homogeneous 405

matrix
– hermitian 9
– probability transition 406
matrix of amplitude 250
mean probability
– error 346, 458
measure
– σ-finite 5
– atomic 72
– discrete 72
– operator-valued 58
– projection-valued 60, 61
– quantum probability 58
– vector 83
measurement 60
– quantum 73
measures
– tight 79
memory channel
– concatenated 390
modulus
– complex 1

net 34
non-expansive 39
norm
– complete bound 149
– diamond 149
– energy-constrained diamond 156
– Hilbert–Schmidt 29
– Hilbertian 2
– operator 10
normal 41
notation
– Dirac 5
number
– natural 2
– Schmidt 244

observable 59
– local 393
– quantum 59
operation
– extended quantum 118
– flip 403
– partial flip 403
– quantum 118
operator
– adjoint 8
– affiliated 41
– bounded linear 10
– closed 8
– compact 22
– contraction 10
– $\mathfrak{H}$- 77
– Hilbert–Schmidt 30
– identity 4, 10
– linear 8
– local 392
– positive 14, 34
– projection 17
– self-adjoint 9
– shift 394
– symmetric 9
– trace class 26
– unitary 16
– zero 10
operators
– direct sum 66
output quantity
– Holevo 296

part
– imaginary 1
– real 1
PCE 239
PFE 239
picture
– Heisenberg 390
– Heisensberg 121
– Schrodinger 119, 385
– Shrodinger 390
point
– extreme 81
polar decomposition 23
positivity
– complete 97
POVM 58
preamble 422, 423, 446
predual 46
probability
– error 356
probability space
– classical 59
– quantum 59, 60
product
– inner 2
– tensor 62

projection
– orthogonal 19
projector
– spectral 77
property
– sufficiency 131
proposition 341
purification 130

QECC 339
quantity
– Holevo 295
– unconstrained χ-function 296
– unconstrained input Holevo 295

Radon–Nikodym derivative 203
range 8
relation
– duality 127
representation 56
– cyclic 57
– faithful 56
– GNS 56
– Kraus 128
– minimal Stinespring 123
– normal 97
– Stinespring 110, 123
– unitary 390
resolution of identity 61
resolvent 15

semicontinuity
– lower 174
seminorm 161
sequence
– approximating 307
– Cauchy 3
– typical 349
set
– resolvent 15
space
– Borel measurable 5
– complex Hilbert 2
– dual 42
– quantum probability 60
– representation 56
– separable 4
spectrum 15
– continuous 15
– point 15
– residual 15
square root 15
state 54
– average 296
– entangled 255
– Gibbs 177
– input 119
– mixed 55
– optimal average 307
– output 119
– pure 33, 54, 55
– quantum 55
– separable 255
states
– separable 263
strong convergence 161
subadditivity 175, 337
– strong 176, 220
subcovering 3
subset
– dense 3
– total 4
subspace
– complementary 17
– final 119
– initial 119
superentanglement 338
support 8, 72
system
– bipartite 250
– classical-quantum 340
– composite 33
– hybrid 340
systems
– composite 67

tensor product 63
Theorem
– Uhlmann's 138
theorem
– Banach–Alaoglu 7
– Banach–Schauder 10
Theorem
– Banach–Steinhaus 43
theorem
– Chapman–Kolmogorov 406
– closed graph 13
– dominated convergence 83, 176

Theorem
– Fredholm Alternative 22
theorem
– Hahn–Banach 43
– Kraus 116
Theorem
– Krein–Millman 81
theorem
– monotone convergence 14
– Petz’s 226
– Prokhorov 79
– Riesz Representation 42
– spectral 23, 61
– Uniform Boundedness 6
– von Neumann commutant 40
theorem, von Neumann double commutant 40
topology
– σ-weak 35, 46
– weak* 35
trace 25
– partial 68
trace non-increasing 117
trace preserving 117
transform
– Fenchel 92
transformation
– linear 7
typicality
– quantum 349

unit
– imaginary 2, 11

value
– singular 25
vector
– cyclic 57
vectors
– orthogonal 4