Theory of Quantum Information

John Watrous

Institute for Quantum Computing
University of Waterloo

Contents

Preface

This is the first complete draft of a book that began as a set of course notes for a graduate course on the theory of quantum information that I have taught several times at the University of Waterloo.

The book is primarily intended for graduate students and researchers having some familiarity with quantum information and computation, such as would be covered in an introductory-level undergraduate or graduate course on the subject. The focus of the book is on the mathematical aspects of quantum information, with an emphasis on proofs. No attention is paid to motives for studying the theory of quantum information, as it is assumed that the reader has already been motivated—and is perhaps interested in proving new theorems on quantum information of his or her own. It should also be said that this is not a physics book: the Schrödinger equation will not be found herein, and the difficult technological challenge of building quantum information processing devices is blissfully ignored.

The selection of topics covered in this book is not intended to be fully representative of the diverse subject of quantum information science. There is, for example, no discussion of quantum cryptography, quantum error correcting codes and fault-tolerance, quantum algorithms and complexity theory, or topological quantum computing, which are among the topics within the theoretical branches of quantum information science having fundamental importance. Nevertheless, one is likely to encounter some of the core mathematical notions discussed in this book when studying these and other topics.

As the students who have taken my course on the theory of quantum information will attest, I sometimes choose to deviate from the standard conventions of quantum information and computation, particularly with respect to notation and terminology. I have exhibited this behavior once again when writing this book. For example, I have avoided the use of the

commonly used Dirac notation, and in some cases I have changed the names and symbols associated with concepts as I have seen fit. I hope that readers who have previously grown familiar with the notation and conventions of quantum information that I have chosen not to follow will excuse me for this, and hope that they will find value in this book nevertheless.

Each chapter aside from the first includes a collection of exercises, some of which can reasonably be viewed as straightforward, and some of which are much more difficult. In some cases, these exercises have been derived from research papers that clearly reveal their solutions, and I have not attempted to disguise this fact or hide their source. While the exercises may potentially be useful to course instructors, their true purpose is to be useful to students of the subject; there is no substitute for the learning experience to be found in wrestling with (and ideally solving) a difficult problem.

As this is a first draft, I expect that there will be many errors, and will be appreciative of errors being brought to my attention. I thank the following people for pointing out errors in my course notes and previous versions of some of the chapters of this book, as well as for their valuable comments and suggestions: Alessandro Cosentino, Mohammad Derakhshani, Edward Effros, Chris Ferrie, Mirmojtaba Gharibi, Gus Gutoski, Anirudh Krishna, Leung Ming Lam, Alexandre Laplante, Abel Molina, Adam Meikle, Maris Ozols, Daniel Puzzuoli, Ansis Rosmanis, Vincent Russo, Yuan Su, Le Phuc Thinh, and Nengkun Yu. I also thank Debbie Leung, Ashwin Nayak, Marco Piani, and Patrick Hayden for helpful discussions on some of the topics that have been covered in this book, and Sascha Agne for assistance with German translations.

The Institute for Quantum Computing and the School of Computer Science at the University of Waterloo have provided me with both the opportunity to write this book and with an environment in which it was possible, for which I am grateful. I am also grateful to the Natural Sciences and Engineering Research Council of Canada and the Canadian Institute for Advanced Research for their financial support of my research program.

Finally, I thank Christiane Lemieux for encouraging my efforts to write this book on too many occasions to count.

John Watrous
john.watrous@uwaterloo.ca
Waterloo, September 2015

Chapter 1

Mathematical preliminaries

This chapter is intended to serve as a review of the mathematical concepts to be used throughout this book, and also as a reference to be consulted as subsequent chapters are studied, if the need should arise. The first section focuses on linear algebra, and the second on analysis and related topics. Unlike the other chapters in this book, the present chapter does not include proofs, and is not intended to serve as a primary source for the material it reviews—a collection of references provided at the end of the chapter may be consulted by readers interested in a proper development of this material.

1.1 Linear algebra

The theory of quantum information relies heavily on linear algebra in finite-dimensional spaces. The subsections that follow present an overview of the aspects of this subject that are most relevant within the theory of quantum information. It will be assumed that the reader is already familiar with the most basic notions of linear algebra, including those of linear dependence and independence, subspaces, spanning sets, bases, and dimension.

1.1.1 Complex Euclidean spaces

The notion of a complex Euclidean space is used throughout this book. One associates a complex Euclidean space with every discrete and finite system; and fundamental notions such as states and measurements of systems are represented in linear-algebraic terms that refer to these spaces.

Definition of complex Euclidean spaces

An *alphabet* is a finite and nonempty set, whose elements may be considered to be *symbols*. Alphabets will generally be denoted by capital Greek letters, including Σ, Γ, and Δ, while lower case Roman letters near the beginning of the alphabet, including a, b, c, and d, will be used to denote symbols in alphabets. Examples of alphabets include the *binary alphabet* $\{0,1\}$, the n-fold Cartesian product $\{0,1\}^n$ of the binary alphabet with itself, and the alphabet $\{1,\ldots,n\}$, for n being a fixed positive integer.

For any alphabet Σ, one denotes by $\mathbb{C}^\Sigma$ the set of all functions from Σ to the complex numbers $\mathbb{C}$. The set $\mathbb{C}^\Sigma$ forms a vector space of dimension $|\Sigma|$ over the complex numbers when addition and scalar multiplication are defined in the following standard way:

1. Addition: for vectors $u, v \in \mathbb{C}^\Sigma$, the vector $u + v \in \mathbb{C}^\Sigma$ is defined by the equation $(u+v)(a) = u(a) + v(a)$ for all $a \in \Sigma$.
2. Scalar multiplication: for a vector $u \in \mathbb{C}^\Sigma$ and a scalar $\alpha \in \mathbb{C}$, the vector $\alpha u \in \mathbb{C}^\Sigma$ is defined by the equation $(\alpha u)(a) = \alpha u(a)$ for all $a \in \Sigma$.

A vector space defined in this way will be called a *complex Euclidean space*.[1] The value $u(a)$ is referred to as the entry of u indexed by a, for each $u \in \mathbb{C}^\Sigma$ and $a \in \Sigma$. The vector whose entries are all zero is simply denoted 0.

Complex Euclidean spaces will be denoted by scripted capital letters near the end of the alphabet, such as $\mathcal{W}$, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. Subsets of these spaces will also be denoted by scripted letters, and when possible this book will follow a convention to use letters such as $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ near the beginning of the alphabet when these subsets are not necessarily vector spaces. Vectors will be denoted by lowercase Roman letters, again near the end of the alphabet, such as u, v, w, x, y, and z.

When n is a positive integer, one typically writes $\mathbb{C}^n$ rather than $\mathbb{C}^{\{1,\ldots,n\}}$, and it is also typical that one views a vector $u \in \mathbb{C}^n$ as an n-tuple of the form $u = (\alpha_1, \ldots, \alpha_n)$, or as a column vector of the form

$$u = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \tag{1.1}$$

[1] Many quantum information theorists prefer to use the term *Hilbert space*. The term *complex Euclidean space* will be preferred in this book, however, as the term *Hilbert space* refers to a more general notion that allows the possibility of infinite index sets.

for complex numbers $\alpha_1, \ldots, \alpha_n$.

For an arbitrary alphabet Σ, the complex Euclidean space $\mathbb{C}^\Sigma$ may be viewed as being equivalent to $\mathbb{C}^n$ for $n = |\Sigma|$; one simply fixes a bijection

$$f : \{1, \ldots, n\} \to \Sigma \tag{1.2}$$

and associates each vector $u \in \mathbb{C}^\Sigma$ with the vector in $\mathbb{C}^n$ whose k-th entry is $u(f(k))$, for each $k \in \{1, \ldots, n\}$. This may be done implicitly when there is a natural or obviously preferred choice for the bijection f. For example, the elements of the alphabet $\Sigma = \{0,1\}^2$ are naturally ordered as follows: 00, 01, 10, 11. Each vector $u \in \mathbb{C}^\Sigma$ may therefore be associated with the 4-tuple

$$(u(00), u(01), u(10), u(11)), \tag{1.3}$$

or with the column vector

$$\begin{pmatrix} u(00) \\ u(01) \\ u(10) \\ u(11) \end{pmatrix}, \tag{1.4}$$

when it is convenient to do this. While little or no generality would be lost in restricting one's attention to complex Euclidean spaces of the form $\mathbb{C}^n$ for this reason, it is both natural and convenient within computational and information-theoretic settings to allow complex Euclidean spaces to be indexed by arbitrary alphabets.

Inner products and norms of vectors

The *inner product* $\langle u, v \rangle$ of two vectors $u, v \in \mathbb{C}^\Sigma$ is defined as

$$\langle u, v \rangle = \sum_{a \in \Sigma} \overline{u(a)}\, v(a). \tag{1.5}$$

It may be verified that the inner product satisfies the following properties:

1. Linearity in the second argument:

$$\langle u, \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle \tag{1.6}$$

for all $u, v, w \in \mathbb{C}^\Sigma$ and $\alpha, \beta \in \mathbb{C}$.

2. Conjugate symmetry:
$$\langle u, v \rangle = \overline{\langle v, u \rangle} \tag{1.7}$$
for all $u, v \in \mathbb{C}^\Sigma$.

3. Positive definiteness:
$$\langle u, u \rangle \geq 0 \tag{1.8}$$
for all $u \in \mathbb{C}^\Sigma$, with equality if and only if $u = 0$.

It is typical that any function satisfying these three properties is referred to as an inner product, but this is the only inner product for vectors in complex Euclidean spaces that is considered in this book.

The *Euclidean norm* of a vector $u \in \mathbb{C}^\Sigma$ is defined as
$$\|u\| = \sqrt{\langle u, u \rangle} = \sqrt{\sum_{a \in \Sigma} |u(a)|^2}. \tag{1.9}$$

The Euclidean norm possesses the following properties, which define the more general notion of a norm:

1. Positive definiteness: $\|u\| \geq 0$ for all $u \in \mathbb{C}^\Sigma$, with $\|u\| = 0$ if and only if $u = 0$.
2. Positive scalability: $\|\alpha u\| = |\alpha| \|u\|$ for all $u \in \mathbb{C}^\Sigma$ and $\alpha \in \mathbb{C}$.
3. The triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in \mathbb{C}^\Sigma$.

The *Cauchy–Schwarz inequality* states that
$$|\langle u, v \rangle| \leq \|u\| \, \|v\| \tag{1.10}$$
for all $u, v \in \mathbb{C}^\Sigma$, with equality if and only if u and v are linearly dependent. The collection of all unit vectors in a complex Euclidean space $\mathcal{X}$ is called the *unit sphere* in that space, and is denoted
$$\mathcal{S}(\mathcal{X}) = \big\{ u \in \mathcal{X} : \|u\| = 1 \big\}. \tag{1.11}$$

The Euclidean norm represents the case $p = 2$ of the class of *p-norms*, defined for each $u \in \mathbb{C}^\Sigma$ as
$$\|u\|_p = \left(\sum_{a \in \Sigma} |u(a)|^p \right)^{\frac{1}{p}} \tag{1.12}$$

for $p < \infty$, and

$$\|u\|_\infty = \max\bigl\{|u(a)| \,:\, a \in \Sigma\bigr\}. \tag{1.13}$$

The above three norm properties (positive definiteness, positive scalability, and the triangle inequality) hold for $\|\cdot\|$ replaced by $\|\cdot\|_p$ for any choice of $p \in [1, \infty]$.

Orthogonality and orthonormality

Two vectors $u, v \in \mathbb{C}^\Sigma$ are said to be *orthogonal* if and only if $\langle u, v\rangle = 0$. The notation $u \perp v$ is also used to indicate that u and v are orthogonal. More generally, for any set $\mathcal{A} \subseteq \mathbb{C}^\Sigma$, the notation $u \perp \mathcal{A}$ indicates that $\langle u, v\rangle = 0$ for all vectors $v \in \mathcal{A}$.

A collection of vectors

$$\{u_a \,:\, a \in \Gamma\} \subset \mathbb{C}^\Sigma, \tag{1.14}$$

indexed by an alphabet Γ, is said to be an *orthogonal set* if it holds that $\langle u_a, u_b\rangle = 0$ for all choices of $a, b \in \Gamma$ with $a \neq b$. A collection of nonzero orthogonal vectors is necessarily linearly independent.

An orthogonal set of *unit* vectors is called an *orthonormal set*, and when such a set forms a basis it is called an *orthonormal basis*. It holds that an orthonormal set

$$\{u_a \,:\, a \in \Gamma\} \subset \mathbb{C}^\Sigma \tag{1.15}$$

is an orthonormal basis of $\mathbb{C}^\Sigma$ if and only if $|\Gamma| = |\Sigma|$. The *standard basis* of $\mathbb{C}^\Sigma$ is the orthonormal basis given by $\{e_a \,:\, a \in \Sigma\}$, where

$$e_a(b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases} \tag{1.16}$$

for all $a, b \in \Sigma$.

Direct sums of complex Euclidean spaces

The *direct sum* of n complex Euclidean spaces $\mathcal{X}_1 = \mathbb{C}^{\Sigma_1}, \ldots, \mathcal{X}_n = \mathbb{C}^{\Sigma_n}$ is the complex Euclidean space

$$\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_n = \mathbb{C}^{\Sigma_1 \sqcup \cdots \sqcup \Sigma_n}, \tag{1.17}$$

where $\Sigma_1 \sqcup \cdots \sqcup \Sigma_n$ denotes the *disjoint union* of the alphabets $\Sigma_1, \ldots, \Sigma_n$, defined as

$$\Sigma_1 \sqcup \cdots \sqcup \Sigma_n = \bigcup_{k \in \{1,\ldots,n\}} \bigl\{(k, a) \,:\, a \in \Sigma_k\bigr\}. \tag{1.18}$$

For vectors $u_1 \in \mathcal{X}_1, \ldots, u_n \in \mathcal{X}_n$, the notation $u_1 \oplus \cdots \oplus u_n \in \mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_n$ refers to the vector for which

$$(u_1 \oplus \cdots \oplus u_n)(k, a) = u_k(a), \tag{1.19}$$

for each $k \in \{1, \ldots, n\}$ and $a \in \Sigma_k$. If each u_k is viewed as a column vector of dimension $|\Sigma_k|$, the vector $u_1 \oplus \cdots \oplus u_n$ may be viewed as a column vector

$$\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \tag{1.20}$$

having dimension $|\Sigma_1| + \cdots + |\Sigma_n|$.

Every element of the space $\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_n$ can be written as $u_1 \oplus \cdots \oplus u_n$ for a unique choice of vectors $u_1, \ldots, u_n$. The following identities hold for every choice of $u_1, v_1 \in \mathcal{X}_1, \ldots, u_n, v_n \in \mathcal{X}_n$, and $\alpha \in \mathbb{C}$:

$$u_1 \oplus \cdots \oplus u_n + v_1 \oplus \cdots \oplus v_n = (u_1 + v_1) \oplus \cdots \oplus (u_n + v_n), \tag{1.21}$$

$$\alpha(u_1 \oplus \cdots \oplus u_n) = (\alpha u_1) \oplus \cdots \oplus (\alpha u_n), \tag{1.22}$$

$$\langle u_1 \oplus \cdots \oplus u_n, v_1 \oplus \cdots \oplus v_n \rangle = \langle u_1, v_1 \rangle + \cdots + \langle u_n, v_n \rangle. \tag{1.23}$$

Tensor products of complex Euclidean spaces

The *tensor product* of n complex Euclidean spaces $\mathcal{X}_1 = \mathbb{C}^{\Sigma_1}, \ldots, \mathcal{X}_n = \mathbb{C}^{\Sigma_n}$ is the complex Euclidean space

$$\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n = \mathbb{C}^{\Sigma_1 \times \cdots \times \Sigma_n}. \tag{1.24}$$

For vectors $u_1 \in \mathcal{X}_1, \ldots, u_n \in \mathcal{X}_n$, the notation $u_1 \otimes \cdots \otimes u_n \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ refers to the vector for which

$$(u_1 \otimes \cdots \otimes u_n)(a_1, \ldots, a_n) = u_1(a_1) \cdots u_n(a_n). \tag{1.25}$$

Vectors of the form $u_1 \otimes \cdots \otimes u_n$ are called *elementary tensors*. They span the space $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, but not every element of $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ is an elementary tensor.

The following identities hold for all vectors $u_1, v_1 \in \mathcal{X}_1, \ldots, u_n, v_n \in \mathcal{X}_n$, scalars $\alpha, \beta \in \mathbb{C}$, and indices $k \in \{1, \ldots, n\}$:

$$\begin{aligned} u_1 \otimes \cdots \otimes u_{k-1} \otimes (\alpha u_k + \beta v_k) \otimes u_{k+1} \otimes \cdots \otimes u_n & \\ = \alpha \left(u_1 \otimes \cdots \otimes u_{k-1} \otimes u_k \otimes u_{k+1} \otimes \cdots \otimes u_n\right) & \\ + \beta \left(u_1 \otimes \cdots \otimes u_{k-1} \otimes v_k \otimes u_{k+1} \otimes \cdots \otimes u_n\right), & \end{aligned} \tag{1.26}$$

$$\langle u_1 \otimes \cdots \otimes u_n, v_1 \otimes \cdots \otimes v_n \rangle = \langle u_1, v_1 \rangle \cdots \langle u_n, v_n \rangle. \tag{1.27}$$

Tensor products are often defined in a way that is more abstract (and more generally applicable) than the definition above, which is sometimes known more specifically as the *Kronecker product*. The following proposition is a reflection of the more abstract definition.

Proposition 1.1. *Let $\mathcal{X}_1, \ldots, \mathcal{X}_n$ and $\mathcal{Y}$ be complex Euclidean spaces and let*

$$\phi : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathcal{Y} \tag{1.28}$$

be a multilinear function, meaning a function for which the mapping

$$u_k \mapsto \phi(u_1, \ldots, u_n) \tag{1.29}$$

is linear for all $k \in \{1, \ldots, n\}$ and all fixed choices of $u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_n$. There exists a unique linear mapping

$$A : \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \to \mathcal{Y} \tag{1.30}$$

such that

$$\phi(u_1, \ldots, u_n) = A(u_1 \otimes \cdots \otimes u_n) \tag{1.31}$$

for all choices of $u_1 \in \mathcal{X}_1, \ldots, u_n \in \mathcal{X}_n$.

If $\mathcal{X}$ is a complex Euclidean space, $x \in \mathcal{X}$ is a vector, and n is a positive integer, then the notations $\mathcal{X}^{\otimes n}$ and $x^{\otimes n}$ refer to the n-fold tensor product of either $\mathcal{X}$ or x with itself.

Real Euclidean spaces

Real Euclidean spaces are defined in a similar way to complex Euclidean spaces, except that the field of complex numbers $\mathbb{C}$ is replaced by the field of real numbers $\mathbb{R}$ in each of the definitions and concepts in which it arises.

Naturally, complex conjugation acts trivially in the real case, and therefore may be omitted.

Complex Euclidean spaces will play a more prominent role than real Euclidean spaces in this book. Real Euclidean spaces will, nevertheless, be particularly important in those settings that make use of concepts from the theory of convexity. The space of Hermitian operators acting on a given complex Euclidean space is an important example of a real vector space that can be identified with a real Euclidean space, as is discussed in the subsection following this one.

1.1.2 Linear operators

Given complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, one writes $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ to refer to the collection of all linear mappings of the form

$$A : \mathcal{X} \to \mathcal{Y}. \tag{1.32}$$

Such mappings will be referred to as *linear operators*, or simply *operators*, from $\mathcal{X}$ to $\mathcal{Y}$ in this book. Parentheses are omitted when expressing the action of linear operators on vectors when no confusion arises in doing so. For instance, one writes Au rather than $A(u)$ to denote the vector resulting from the application of an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ to a vector $u \in \mathcal{X}$.

The set $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ forms a complex vector space when addition and scalar multiplication are defined as follows:

1. Addition: for operators $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, the operator $A + B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ is defined by the equation

   $$(A + B)u = Au + Bu \tag{1.33}$$

   for all $u \in \mathcal{X}$.

2. Scalar multiplication: for an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and a scalar $\alpha \in \mathbb{C}$, the operator $\alpha A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ is defined by the equation

   $$(\alpha A)u = \alpha\, Au \tag{1.34}$$

   for all $u \in \mathcal{X}$.

Matrices and their correspondence with operators

A *matrix* over the complex numbers is a mapping of the form

$$M : \Gamma \times \Sigma \to \mathbb{C} \tag{1.35}$$

for alphabets Σ and Γ. For $a \in \Gamma$ and $b \in \Sigma$ the value $M(a, b)$ is called the (a, b) *entry* of M, and the elements a and b are referred to as *indices* in this context: a is the *row index* and b is the *column index* of the entry $M(a, b)$. Addition and scalar multiplication of matrices are defined in a similar way to vectors in complex Euclidean spaces:

1. Addition: for matrices $M : \Gamma \times \Sigma \to \mathbb{C}$ and $N : \Gamma \times \Sigma \to \mathbb{C}$, the matrix $M + N$ is defined as

   $$(M + N)(a, b) = M(a, b) + N(a, b) \tag{1.36}$$

   for all $a \in \Gamma$ and $b \in \Sigma$.

2. Scalar multiplication: for a matrix $M : \Gamma \times \Sigma \to \mathbb{C}$ and a scalar $\alpha \in \mathbb{C}$, the matrix αM is defined as

   $$(\alpha M)(a, b) = \alpha M(a, b) \tag{1.37}$$

   for all $a \in \Gamma$ and $b \in \Sigma$.

In addition, one defines matrix multiplication as follows:

3. Matrix multiplication: for matrices $M : \Gamma \times \Delta \to \mathbb{C}$ and $N : \Delta \times \Sigma \to \mathbb{C}$, the matrix $MN : \Gamma \times \Sigma \to \mathbb{C}$ is defined as

   $$(MN)(a, b) = \sum_{c \in \Delta} M(a, c) N(c, b) \tag{1.38}$$

   for all $a \in \Gamma$ and $b \in \Sigma$.

For any choice of complex Euclidean spaces $\mathcal{X} = \mathbb{C}^{\Sigma}$ and $\mathcal{Y} = \mathbb{C}^{\Gamma}$, there is a bijective linear correspondence between the set of operators $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ and the collection of all matrices taking the form $M : \Gamma \times \Sigma \to \mathbb{C}$ that is obtained as follows. With each operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, one associates the matrix M defined as

$$M(a, b) = \langle e_a, A e_b \rangle \tag{1.39}$$

for $a \in \Gamma$ and $b \in \Sigma$. The operator A is uniquely determined by M, and may be recovered from M by the equation

$$(Au)(a) = \sum_{b \in \Sigma} M(a,b)u(b) \tag{1.40}$$

for all $a \in \Gamma$. With respect to this correspondence, matrix multiplication is equivalent to operator composition.

Hereafter in this book, linear operators will be associated with matrices implicitly, without the introduction of names that distinguish matrices from the operators they are associated with. With this in mind, the notation

$$A(a,b) = \langle e_a, Ae_b \rangle \tag{1.41}$$

is introduced for each $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, $a \in \Gamma$, and $b \in \Sigma$ (where it is to be assumed that $\mathcal{X} = \mathbb{C}^\Sigma$ and $\mathcal{Y} = \mathbb{C}^\Gamma$ as before).

The standard basis of a space of operators

For every choice of complex Euclidean spaces $\mathcal{X} = \mathbb{C}^\Sigma$ and $\mathcal{Y} = \mathbb{C}^\Gamma$, and each choice of symbols $a \in \Gamma$ and $b \in \Sigma$, the operator $E_{a,b} \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ is defined as

$$E_{a,b}\, u = u(b)e_a \tag{1.42}$$

for every $u \in \mathcal{X}$. Equivalently, $E_{a,b}$ is defined by the equation

$$E_{a,b}(c,d) = \begin{cases} 1 & \text{if } (c,d) = (a,b) \\ 0 & \text{otherwise} \end{cases} \tag{1.43}$$

holding for all $c \in \Gamma$ and $d \in \Sigma$. The collection

$$\{E_{a,b} \,:\, a \in \Gamma,\ b \in \Sigma\} \tag{1.44}$$

forms a basis of $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ known as the *standard basis* of this space. The number of elements in this basis is, of course, consistent with the fact that the dimension of $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ is given by $\dim(\mathrm{L}(\mathcal{X}, \mathcal{Y})) = \dim(\mathcal{X}) \dim(\mathcal{Y})$.

The entry-wise conjugate, transpose, and adjoint

For every operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, for complex Euclidean spaces $\mathcal{X} = \mathbb{C}^\Sigma$ and $\mathcal{Y} = \mathbb{C}^\Gamma$, one defines three additional operators,

$$\overline{A} \in \mathrm{L}(\mathcal{X}, \mathcal{Y}) \quad \text{and} \quad A^{\mathsf{T}}, A^* \in \mathrm{L}(\mathcal{Y}, \mathcal{X}), \tag{1.45}$$

as follows:

1. The operator $\overline{A} \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ is the operator whose matrix representation has entries that are complex conjugates to the matrix representation of A:
$$\overline{A}(a, b) = \overline{A(a, b)} \tag{1.46}$$
for all $a \in \Gamma$ and $b \in \Sigma$.

2. The operator $A^{\mathsf{T}} \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ is the operator whose matrix representation is obtained by *transposing* the matrix representation of A:
$$A^{\mathsf{T}}(b, a) = A(a, b) \tag{1.47}$$
for all $a \in \Gamma$ and $b \in \Sigma$.

3. The operator $A^* \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ is the uniquely determined operator that satisfies the equation
$$\langle v, Au \rangle = \langle A^* v, u \rangle \tag{1.48}$$
for all $u \in \mathcal{X}$ and $v \in \mathcal{Y}$. It may be obtained by performing both of the operations described in items 1 and 2:
$$A^* = \overline{A^{\mathsf{T}}}. \tag{1.49}$$

The operators $\overline{A}$, A^{T}, and A^* will be called the *entry-wise conjugate*, *transpose*, and *adjoint* operators to A, respectively.

The mappings $A \mapsto \overline{A}$ and $A \mapsto A^*$ are conjugate linear and $A \mapsto A^{\mathsf{T}}$ is linear:
$$\begin{aligned}
\overline{\alpha A + \beta B} &= \overline{\alpha}\,\overline{A} + \overline{\beta}\,\overline{B}, \\
(\alpha A + \beta B)^* &= \overline{\alpha} A^* + \overline{\beta} B^*, \\
(\alpha A + \beta B)^{\mathsf{T}} &= \alpha A^{\mathsf{T}} + \beta B^{\mathsf{T}},
\end{aligned}$$
for all $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and $\alpha, \beta \in \mathbb{C}$. These mappings are bijections, each being its own inverse.

Each vector $u \in \mathcal{X}$ in a complex Euclidean space $\mathcal{X}$ may be identified with the linear operator in $\mathrm{L}(\mathbb{C}, \mathcal{X})$ defined as $\alpha \mapsto \alpha u$ for all $\alpha \in \mathbb{C}$. Through this identification, the linear mappings $\overline{u} \in \mathrm{L}(\mathbb{C}, \mathcal{X})$ and $u^{\mathsf{T}}, u^* \in \mathrm{L}(\mathcal{X}, \mathbb{C})$ are defined as above. As an element of $\mathcal{X}$, the vector $\overline{u}$ is simply the entry-wise complex conjugate of u, i.e., if $\mathcal{X} = \mathbb{C}^{\Sigma}$ then
$$\overline{u}(a) = \overline{u(a)} \tag{1.50}$$
for every $a \in \Sigma$. For each vector $u \in \mathcal{X}$ the mapping $u^* \in \mathrm{L}(\mathcal{X}, \mathbb{C})$ satisfies $u^* v = \langle u, v \rangle$ for all $v \in \mathcal{X}$.

Kernel, image, and rank

The *kernel* of an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ is the subspace of $\mathcal{X}$ defined as

$$\ker(A) = \{u \in \mathcal{X} \,:\, Au = 0\}, \tag{1.51}$$

while the *image* of A is the subspace of $\mathcal{Y}$ defined as

$$\mathrm{im}(A) = \{Au \,:\, u \in \mathcal{X}\}. \tag{1.52}$$

For every operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, one has that

$$\ker(A) = \ker(A^*A) \quad \text{and} \quad \mathrm{im}(A) = \mathrm{im}(AA^*), \tag{1.53}$$

as well as the equation

$$\dim(\ker(A)) + \dim(\mathrm{im}(A)) = \dim(\mathcal{X}). \tag{1.54}$$

The *rank* of an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, denoted $\mathrm{rank}(A)$, is the dimension of the image of A:

$$\mathrm{rank}(A) = \dim(\mathrm{im}(A)). \tag{1.55}$$

By the second equation of (1.53) it follows that $\mathrm{rank}(A) = \mathrm{rank}(AA^*)$ for every $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$.

For any choice of vectors $u \in \mathcal{X}$ and $v \in \mathcal{Y}$, the operator $vu^* \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ satisfies

$$(vu^*)w = v(u^*w) = \langle u, w\rangle v \tag{1.56}$$

for all $w \in \mathcal{X}$. Assuming that u and v are nonzero, the operator vu^* has rank equal to one, and every rank one operator in $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ can be expressed in this form for vectors u and v that are unique up to scalar multiples.

Operators involving direct sums of complex Euclidean spaces

Suppose that $\mathcal{X}_1 = \mathbb{C}^{\Sigma_1}, \ldots, \mathcal{X}_n = \mathbb{C}^{\Sigma_n}$ and $\mathcal{Y}_1 = \mathbb{C}^{\Gamma_1}, \ldots, \mathcal{Y}_m = \mathbb{C}^{\Gamma_m}$ are complex Euclidean spaces, for any choice of alphabets $\Sigma_1, \ldots, \Sigma_n$ and $\Gamma_1, \ldots, \Gamma_m$. For a given operator

$$A \in \mathrm{L}(\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_n, \mathcal{Y}_1 \oplus \cdots \oplus \mathcal{Y}_m), \tag{1.57}$$

there exists a unique collection of operators

$$\bigl\{A_{j,k} \in \mathrm{L}(\mathcal{X}_k, \mathcal{Y}_j) \,:\, 1 \leq j \leq m,\, 1 \leq k \leq n\bigr\} \tag{1.58}$$

for which the equation

$$A_{j,k}(a,b) = A\big((j,a),(k,b)\big) \tag{1.59}$$

holds for all $j \in \{1,\ldots,m\}$, $k \in \{1,\ldots,n\}$, $a \in \Gamma_j$, and $b \in \Sigma_k$. For all vectors $u_1 \in \mathcal{X}_1, \ldots, u_n \in \mathcal{X}_n$, one has that

$$A(u_1 \oplus \cdots \oplus u_n) = v_1 \oplus \cdots \oplus v_m \tag{1.60}$$

for $v_1 \in \mathcal{Y}_1, \ldots, v_m \in \mathcal{Y}_m$ being defined as

$$v_j = \sum_{k=1}^{n} A_{j,k} u_k \tag{1.61}$$

for each $j \in \{1,\ldots,m\}$. Conversely, for any collection of operators of the form (1.58), there is a unique operator A of the form (1.57) that obeys the equations (1.60) and (1.61) for all vectors $u_1 \in \mathcal{X}_1, \ldots, u_n \in \mathcal{X}_n$.

There is therefore a bijective correspondence between operators of the form (1.57) and collections of operators of the form (1.58). With respect to the matrix representations of these operators, this correspondence may be expressed succinctly as

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix}. \tag{1.62}$$

One interprets the right-hand side of (1.62) as the specification of the operator having the form (1.57) that is defined by the collection (1.58) in this way.

Tensor products of operators

Suppose that $\mathcal{X}_1 = \mathbb{C}^{\Sigma_1}, \ldots, \mathcal{X}_n = \mathbb{C}^{\Sigma_n}$ and $\mathcal{Y}_1 = \mathbb{C}^{\Gamma_1}, \ldots, \mathcal{Y}_n = \mathbb{C}^{\Gamma_n}$ are complex Euclidean spaces, for alphabets $\Sigma_1, \ldots, \Sigma_n$ and $\Gamma_1, \ldots, \Gamma_n$. For any choice of operators

$$A_1 \in \mathrm{L}(\mathcal{X}_1, \mathcal{Y}_1), \ \ldots, \ A_n \in \mathrm{L}(\mathcal{X}_n, \mathcal{Y}_n), \tag{1.63}$$

one defines the tensor product

$$A_1 \otimes \cdots \otimes A_n \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n) \tag{1.64}$$

of these operators to be the unique operator that satisfies the equation

$$(A_1 \otimes \cdots \otimes A_n)(u_1 \otimes \cdots \otimes u_n) = (A_1 u_1) \otimes \cdots \otimes (A_n u_n) \tag{1.65}$$

for all choices of $u_1 \in \mathcal{X}_1, \ldots, u_n \in \mathcal{X}_n$. This operator may equivalently be defined in terms of its matrix representation as

$$\begin{aligned}(A_1 \otimes \cdots \otimes A_n)((a_1, \ldots, a_n), (b_1, \ldots, b_n)) \\ = A_1(a_1, b_1) \cdots A_n(a_n, b_n)\end{aligned} \tag{1.66}$$

for all $a_1 \in \Gamma_1, \ldots, a_n \in \Gamma_n$ and $b_1 \in \Sigma_1, \ldots, b_n \in \Sigma_n$.

For every choice of complex Euclidean spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$, $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$, and $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$, operators

$$\begin{aligned}A_1, B_1 \in \mathrm{L}(\mathcal{X}_1, \mathcal{Y}_1), \ \ldots, \ A_n, B_n \in \mathrm{L}(\mathcal{X}_n, \mathcal{Y}_n), \\ C_1 \in \mathrm{L}(\mathcal{Y}_1, \mathcal{Z}_1), \ \ldots, \ C_n \in \mathrm{L}(\mathcal{Y}_n, \mathcal{Z}_n),\end{aligned} \tag{1.67}$$

and scalars $\alpha, \beta \in \mathbb{C}$, the following equations hold:

$$\begin{aligned}A_1 \otimes \cdots \otimes A_{k-1} \otimes (\alpha A_k + \beta B_k) \otimes A_{k+1} \otimes \cdots \otimes A_n \\ = \alpha(A_1 \otimes \cdots \otimes A_{k-1} \otimes A_k \otimes A_{k+1} \otimes \cdots \otimes A_n) \\ + \beta(A_1 \otimes \cdots \otimes A_{k-1} \otimes B_k \otimes A_{k+1} \otimes \cdots \otimes A_n),\end{aligned} \tag{1.68}$$

$$(C_1 \otimes \cdots \otimes C_n)(A_1 \otimes \cdots \otimes A_n) = (C_1 A_1) \otimes \cdots \otimes (C_n A_n), \tag{1.69}$$

$$(A_1 \otimes \cdots \otimes A_n)^{\mathsf{T}} = A_1^{\mathsf{T}} \otimes \cdots \otimes A_n^{\mathsf{T}}, \tag{1.70}$$

$$\overline{A_1 \otimes \cdots \otimes A_n} = \overline{A_1} \otimes \cdots \otimes \overline{A_n}, \tag{1.71}$$

$$(A_1 \otimes \cdots \otimes A_n)^* = A_1^* \otimes \cdots \otimes A_n^*. \tag{1.72}$$

Similar to vectors, for an operator A and a positive integer n, the notation $A^{\otimes n}$ refers to the n-fold tensor product of A with itself.

Square operators

For every complex Euclidean space $\mathcal{X}$, the notation $\mathrm{L}(\mathcal{X})$ is understood to be a shorthand for $\mathrm{L}(\mathcal{X}, \mathcal{X})$. Operators in the space $\mathrm{L}(\mathcal{X})$ will be called *square operators*, due to the fact that their matrix representations are square, with rows and columns indexed by the same set.

The space $\mathrm{L}(\mathcal{X})$ is an *associative algebra*; in addition to being a vector space, the composition of square operators is an associative and bilinear operation:

$$\begin{aligned} (XY)Z &= X(YZ), \\ Z(\alpha X + \beta Y) &= \alpha ZX + \beta ZY, \\ (\alpha X + \beta Y)Z &= \alpha XZ + \beta YZ, \end{aligned} \tag{1.73}$$

for every choice of $X, Y, Z \in \mathrm{L}(\mathcal{X})$ and $\alpha, \beta \in \mathbb{C}$.

The *identity operator* $\mathbb{1} \in \mathrm{L}(\mathcal{X})$ is the operator defined as $\mathbb{1}u = u$ for all $u \in \mathcal{X}$. It may also be defined by its matrix representation as

$$\mathbb{1}(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases} \tag{1.74}$$

for all $a, b \in \Sigma$, assuming $\mathcal{X} = \mathbb{C}^{\Sigma}$. One writes $\mathbb{1}_{\mathcal{X}}$ rather than $\mathbb{1}$ when it is helpful to indicate explicitly that this operator acts on $\mathcal{X}$.

For complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ is *invertible* if there exists an operator $B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ such that $BA = \mathbb{1}_{\mathcal{X}}$. When such an operator B exists it is necessarily unique, and is denoted A^{-1}.

Trace and determinant

The *diagonal* entries of a square operator $X \in \mathrm{L}(\mathcal{X})$, for $\mathcal{X} = \mathbb{C}^{\Sigma}$, are those of the form $X(a, a)$ for $a \in \Sigma$. The *trace* of a square operator $X \in \mathrm{L}(\mathcal{X})$ is defined as the sum of its diagonal entries:

$$\mathrm{Tr}(X) = \sum_{a \in \Sigma} X(a, a). \tag{1.75}$$

Alternatively, the trace is the unique linear function $\mathrm{Tr} : \mathrm{L}(\mathcal{X}) \to \mathbb{C}$ for which it holds that

$$\mathrm{Tr}\big(uv^*\big) = \langle v, u \rangle \tag{1.76}$$

for all vectors $u, v \in \mathcal{X}$.

For every choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$ and operators $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and $B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$, one has

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA). \tag{1.77}$$

This property is known as the *cyclic property* of the trace.

By means of the trace, one defines an inner product on the space $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ as follows:

$$\langle A, B \rangle = \mathrm{Tr}\big(A^* B\big) \tag{1.78}$$

for all $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$. It may be verified that this inner product satisfies the requisite properties of being an inner product:

1. Linearity in the second argument:

$$\langle A, \alpha B + \beta C \rangle = \alpha \langle A, B \rangle + \beta \langle A, C \rangle \tag{1.79}$$

   for all $A, B, C \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and $\alpha, \beta \in \mathbb{C}$.

2. Conjugate symmetry:

$$\langle A, B \rangle = \overline{\langle B, A \rangle} \tag{1.80}$$

   for all $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$.

3. Positive definiteness: $\langle A, A \rangle \geq 0$ for all $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, with equality if and only if $A = 0$.

The *determinant* of a square operator $X \in \mathrm{L}(\mathcal{X})$, for $\mathcal{X} = \mathbb{C}^{\Sigma}$, is defined by the equation

$$\mathrm{Det}(X) = \sum_{\pi \in \mathrm{Sym}(\Sigma)} \mathrm{sign}(\pi) \prod_{a \in \Sigma} X(a, \pi(a)). \tag{1.81}$$

Here, the set $\mathrm{Sym}(\Sigma)$ denotes the collection of all permutations $\pi : \Sigma \to \Sigma$, and $\mathrm{sign}(\pi) \in \{-1, +1\}$ denotes the sign (or parity) of the permutation π. The determinant is multiplicative,

$$\mathrm{Det}(XY) = \mathrm{Det}(X)\,\mathrm{Det}(Y) \tag{1.82}$$

for all $X, Y \in \mathrm{L}(\mathcal{X})$, and $\mathrm{Det}(X) \neq 0$ if and only if X is invertible.

Eigenvectors and eigenvalues

If $X \in \mathrm{L}(\mathcal{X})$ is an operator and $u \in \mathcal{X}$ is a nonzero vector for which it holds that

$$Xu = \lambda u \tag{1.83}$$

for some choice of $\lambda \in \mathbb{C}$, then u is said to be an *eigenvector* of X and λ is its corresponding *eigenvalue*.

For every operator $X \in \mathrm{L}(\mathcal{X})$, one has that

$$p_X(\alpha) = \mathrm{Det}(\alpha \mathbb{1}_{\mathcal{X}} - X) \tag{1.84}$$

is a monic polynomial in the variable α having degree $\dim(\mathcal{X})$, known as the *characteristic polynomial* of X. The *spectrum* of X, denoted $\mathrm{spec}(X)$, is the multiset containing the roots of the polynomial p_X, where each root appears a number of times equal to its multiplicity. As p_X is monic, it holds that

$$p_X(\alpha) = \prod_{\lambda \in \mathrm{spec}(X)} (\alpha - \lambda). \tag{1.85}$$

Each element $\lambda \in \mathrm{spec}(X)$ is necessarily an eigenvalue of X, and every eigenvalue of X is contained in $\mathrm{spec}(X)$.

The trace and determinant may be expressed in terms of the spectrum as follows:

$$\mathrm{Tr}(X) = \sum_{\lambda \in \mathrm{spec}(X)} \lambda \quad \text{and} \quad \mathrm{Det}(X) = \prod_{\lambda \in \mathrm{spec}(X)} \lambda \tag{1.86}$$

for every $X \in \mathrm{L}(\mathcal{X})$. The *spectral radius* of an operator $X \in \mathrm{L}(\mathcal{X})$ is the maximum absolute value $|\lambda|$ taken over all eigenvalues λ of X. For every choice of operators $X, Y \in \mathrm{L}(\mathcal{X})$ it holds that $\mathrm{spec}(XY) = \mathrm{spec}(YX)$.

Lie brackets and commutants

A set $\mathcal{A} \subseteq \mathrm{L}(\mathcal{X})$ is a *subalgebra* of $\mathrm{L}(\mathcal{X})$ if it is closed under addition, scalar multiplication, and operator composition:

$$X + Y \in \mathcal{A}, \quad \alpha X \in \mathcal{A}, \quad \text{and} \quad XY \in \mathcal{A} \tag{1.87}$$

for all $X, Y \in \mathcal{A}$ and $\alpha \in \mathbb{C}$. A subalgebra $\mathcal{A}$ of $\mathrm{L}(\mathcal{X})$ is said to be *self-adjoint* if it holds that $X^* \in \mathcal{A}$ for every $X \in \mathcal{A}$, and is said to be *unital* if it holds that $\mathbb{1} \in \mathcal{A}$.

For any pair of operators $X, Y \in \mathrm{L}(\mathcal{X})$, the *Lie bracket* $[X, Y] \in \mathrm{L}(\mathcal{X})$ is defined as

$$[X, Y] = XY - YX. \tag{1.88}$$

It holds that $[X, Y] = 0$ if and only if X and Y *commute*: $XY = YX$. For any subset of operators $\mathcal{A} \subseteq \mathrm{L}(\mathcal{X})$, one defines the *commutant* of $\mathcal{A}$ as

$$\mathrm{comm}(\mathcal{A}) = \big\{ Y \in \mathrm{L}(\mathcal{X}) \,:\, [X, Y] = 0 \text{ for all } X \in \mathcal{A} \big\}. \tag{1.89}$$

The commutant of every subset of $\mathrm{L}(\mathcal{X})$ is a unital subalgebra of $\mathrm{L}(\mathcal{X})$.

Important classes of operators

The following classes of operators have particular importance in the theory of quantum information.

1. *Normal operators.* An operator $X \in \mathrm{L}(\mathcal{X})$ is a *normal* operator if and only if it commutes with its adjoint: $[X, X^*] = 0$, or equivalently $XX^* = X^*X$. The importance of this collection of operators, for the purposes of this book, is mainly derived from two facts: (1) the normal operators are those for which the spectral theorem (discussed later in Section 1.1.3) holds, and (2) most of the special classes of operators that are discussed below are subsets of the normal operators.

2. *Hermitian operators.* An operator $X \in \mathrm{L}(\mathcal{X})$ is *Hermitian* if and only if $X = X^*$. The set of Hermitian operators acting on a complex Euclidean space $\mathcal{X}$ will hereafter be denoted $\mathrm{Herm}(\mathcal{X})$ in this book:

$$\mathrm{Herm}(\mathcal{X}) = \{X \in \mathrm{L}(\mathcal{X}) \,:\, X = X^*\}. \tag{1.90}$$

   Every Hermitian operator is a normal operator.

3. *Positive semidefinite operators.* An operator $X \in \mathrm{L}(\mathcal{X})$ is *positive semidefinite* if and only if it holds that $X = Y^*Y$ for some operator $Y \in \mathrm{L}(\mathcal{X})$. Positive semidefinite operators will often, as a convention, be denoted by the letters P, Q, and R in this book. The collection of positive semidefinite operators acting on $\mathcal{X}$ is denoted $\mathrm{Pos}(\mathcal{X})$, so that

$$\mathrm{Pos}(\mathcal{X}) = \{Y^*Y \,:\, Y \in \mathrm{L}(\mathcal{X})\}. \tag{1.91}$$

   Every positive semidefinite operator is Hermitian.

4. *Positive definite operators.* A positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X})$ is said to be *positive definite* if, in addition to being positive semidefinite, it is invertible. The notation

$$\mathrm{Pd}(\mathcal{X}) = \{P \in \mathrm{Pos}(\mathcal{X}) \,:\, \mathrm{Det}(P) \neq 0\} \tag{1.92}$$

   will be used to denote the set of such operators for a complex Euclidean space $\mathcal{X}$.

5. *Density operators.* Positive semidefinite operators having trace equal to 1 are called *density operators*. Lowercase Greek letters, such as ρ, ξ, and σ, are conventionally used to denote density operators. The notation

$$\mathrm{D}(\mathcal{X}) = \{\rho \in \mathrm{Pos}(\mathcal{X}) \,:\, \mathrm{Tr}(\rho) = 1\} \tag{1.93}$$

will be used to denote the collection of density operators acting on a complex Euclidean space $\mathcal{X}$.

6. *Projection operators.* A positive semidefinite operator $\Pi \in \mathrm{Pos}(\mathcal{X})$ is said to be a *projection operator*[2] if, in addition to being positive semidefinite, it satisfies the equation $\Pi^2 = \Pi$. Equivalently, a projection operator is a Hermitian operator whose only eigenvalues are 0 and 1. The collection of all projection operators of the form $\Pi \in \mathrm{Pos}(\mathcal{X})$ is denoted $\mathrm{Proj}(\mathcal{X})$. For each subspace $\mathcal{V} \subseteq \mathcal{X}$, there is a uniquely defined projection operator $\Pi \in \mathrm{Proj}(\mathcal{X})$ satisfying $\mathrm{im}(\Pi) = \mathcal{V}$; when it is convenient, the notation $\Pi_{\mathcal{V}}$ is used to refer to this projection operator.

7. *Linear isometries.* An operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ is a *linear isometry* (or simply an *isometry*, for short) if and only if it preserves the Euclidean norm: $\|Au\| = \|u\|$ for all $u \in \mathcal{X}$. The condition that $\|Au\| = \|u\|$ for all $u \in \mathcal{X}$ is equivalent to $A^*A = \mathbb{1}_{\mathcal{X}}$. The notation

$$\mathrm{U}(\mathcal{X}, \mathcal{Y}) = \left\{ A \in \mathrm{L}(\mathcal{X}, \mathcal{Y}) \,:\, A^*A = \mathbb{1}_{\mathcal{X}} \right\} \tag{1.94}$$

   is used to denote this class of operators. In order for a linear isometry of the form $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$ to exist, it must hold that $\dim(\mathcal{Y}) \geq \dim(\mathcal{X})$. Every linear isometry preserves not only the Euclidean norm, but inner products as well: $\langle Au, Av \rangle = \langle u, v \rangle$ for all $u, v \in \mathcal{X}$.

8. *Unitary operators.* The set of isometries mapping a complex Euclidean space $\mathcal{X}$ to itself is denoted $\mathrm{U}(\mathcal{X})$, and operators in this set are *unitary operators*. The letters U, V, and W will often be used to refer to unitary operators (and sometimes to linear isometries more generally) in this book. Every unitary operator $U \in \mathrm{U}(\mathcal{X})$ is necessarily invertible and satisfies the equation $UU^* = U^*U = \mathbb{1}_{\mathcal{X}}$, and is therefore normal.

9. *Diagonal operators.* An operator $X \in \mathrm{L}(\mathcal{X})$, for a complex Euclidean space of the form $\mathcal{X} = \mathbb{C}^{\Sigma}$, is a *diagonal operator* if it holds that $X(a, b) = 0$ for all $a, b \in \Sigma$ with $a \neq b$. For a given vector $u \in \mathcal{X}$, one writes $\mathrm{Diag}(u) \in \mathrm{L}(\mathcal{X})$ to denote the diagonal operator defined as

$$\mathrm{Diag}(u)(a, b) = \begin{cases} u(a) & \text{if } a = b \\ 0 & \text{if } a \neq b. \end{cases} \tag{1.95}$$

[2] Sometimes the term *projection operator* refers to an operator $X \in \mathrm{L}(\mathcal{X})$ that satisfies the equation $X^2 = X$, but that might not be Hermitian. This is not the meaning that is associated with this term in this book.

Further remarks on Hermitian and positive semidefinite operators

The sum of two Hermitian operators is Hermitian, as is a real scalar multiple of a Hermitian operator. The inner product of two Hermitian operators is real as well. For every choice of a complex Euclidean space $\mathcal{X}$, the space $\mathrm{Herm}(\mathcal{X})$ therefore forms a vector space over the real numbers on which an inner product is defined.

Indeed, under the assumption that $\mathcal{X} = \mathbb{C}^{\Sigma}$, it holds that the space $\mathrm{Herm}(\mathcal{X})$ and the real Euclidean space $\mathbb{R}^{\Sigma\times\Sigma}$ are *isometrically isomorphic*: there exists a linear bijection

$$\phi : \mathbb{R}^{\Sigma\times\Sigma} \to \mathrm{Herm}(\mathcal{X}) \tag{1.96}$$

with the property that

$$\langle \phi(u), \phi(v)\rangle = \langle u, v\rangle \tag{1.97}$$

for all $u, v \in \mathbb{R}^{\Sigma\times\Sigma}$. The existence of such a linear bijection allows one to directly translate many statements about real Euclidean spaces to the space of Hermitian operators acting on a complex Euclidean space.

One way to define a mapping ϕ as above is as follows. First, assume that a total ordering of Σ has been fixed, and define a collection

$$\{H_{a,b} \,:\, (a,b) \in \Sigma\times\Sigma\} \subset \mathrm{Herm}(\mathcal{X}) \tag{1.98}$$

as

$$H_{a,b} = \begin{cases} E_{a,a} & \text{if } a = b \\ \frac{1}{\sqrt{2}}(E_{a,b} + E_{b,a}) & \text{if } a < b \\ \frac{1}{\sqrt{2}}(iE_{a,b} - iE_{b,a}) & \text{if } a > b \end{cases} \tag{1.99}$$

for each pair $(a,b) \in \Sigma\times\Sigma$. It holds that (1.98) is an orthonormal set (with respect to the usual inner product defined on $\mathrm{L}(\mathcal{X})$), and moreover every element of $\mathrm{Herm}(\mathcal{X})$ can be expressed uniquely as a real linear combination of the operators in this set. The mapping ϕ defined by the equation

$$\phi\big(e_{(a,b)}\big) = H_{a,b}, \tag{1.100}$$

and extended to all of $\mathbb{R}^{\Sigma\times\Sigma}$ by linearity, satisfies the requirement (1.97).

The eigenvalues of a Hermitian operator are necessarily real numbers, and can therefore be ordered from largest to smallest. For every complex Euclidean space $\mathcal{X}$ and every Hermitian operator $H \in \mathrm{Herm}(\mathcal{X})$, the vector

$$\lambda(H) = (\lambda_1(H),\, \lambda_2(H),\, \ldots,\, \lambda_n(H)) \in \mathbb{R}^n \tag{1.101}$$

is defined so that

$$\operatorname{spec}(H) = \big\{\lambda_1(H),\, \lambda_2(H),\, \ldots,\, \lambda_n(H)\big\} \tag{1.102}$$

and

$$\lambda_1(H) \geq \lambda_2(H) \geq \cdots \geq \lambda_n(H). \tag{1.103}$$

The notation $\lambda_k(H)$ may also be used in isolation to refer to the k-th largest eigenvalue of a Hermitian operator H when necessary. The eigenvalues of Hermitian operators can be characterized by a theorem known as the *Courant–Fischer theorem*, which is as follows.

Theorem 1.2 (Courant–Fischer theorem)**.** *Let $H \in \operatorname{Herm}(\mathcal{X})$ be a Hermitian operator, for $\mathcal{X}$ being a complex Euclidean space of dimension n. For every index $k \in \{1, \ldots, n\}$ it holds that*

$$\begin{aligned} \lambda_k(H) &= \max_{u_1,\ldots,u_{n-k} \in \mathcal{S}(\mathcal{X})} \; \min_{\substack{v \in \mathcal{S}(\mathcal{X}) \\ v \perp \{u_1,\ldots,u_{n-k}\}}} v^* H v \\ &= \min_{u_1,\ldots,u_{k-1} \in \mathcal{S}(\mathcal{X})} \; \max_{\substack{v \in \mathcal{S}(\mathcal{X}) \\ v \perp \{u_1,\ldots,u_{k-1}\}}} v^* H v \end{aligned} \tag{1.104}$$

(It is to be interpreted that the maximum or minimum is simply omitted if it is to be taken over an empty set of vectors, and that $v \perp \varnothing$ holds for all vectors v.)

There are alternative ways to describe positive semidefinite operators that are useful in different situations. In particular, the following statements are equivalent for every operator $P \in \mathrm{L}(\mathcal{X})$:

1. P is positive semidefinite.
2. $P = A^*A$ for an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, for some choice of a complex Euclidean space $\mathcal{Y}$.
3. P is Hermitian and every eigenvalue of P is nonnegative.
4. $\langle u, Pu \rangle$ is a nonnegative real number for every choice of $u \in \mathcal{X}$.
5. $\langle Q, P \rangle$ is a nonnegative real number for every positive semidefinite operator $Q \in \operatorname{Pos}(\mathcal{X})$.
6. There exists a collection of vectors $\{u_a : a \in \Sigma\} \subset \mathcal{X}$ for which it holds that $P(a, b) = \langle u_a, u_b \rangle$ for all $a, b \in \Sigma$.

7. There exists a collection of vectors $\{u_a : a \in \Sigma\} \subset \mathcal{Y}$, for some choice of a complex Euclidean space $\mathcal{Y}$, for which it holds that $P(a,b) = \langle u_a, u_b \rangle$ for all $a, b \in \Sigma$.

Along similar lines, one has that the following statements are equivalent for a given operator $P \in \mathrm{L}(\mathcal{X})$:

1. P is positive definite.
2. P is Hermitian, and every eigenvalue of P is positive.
3. $\langle u, Pu \rangle$ is a positive real number for every nonzero vector $u \in \mathcal{X}$.
4. $\langle Q, P \rangle$ is a positive real number for every nonzero positive semidefinite operator $Q \in \mathrm{Pos}(\mathcal{X})$.
5. There exists a positive real number $\varepsilon > 0$ such that $P - \varepsilon \mathbb{1} \in \mathrm{Pos}(\mathcal{X})$.

The notations $P \geq 0$ and $0 \leq P$ indicate that P is positive semidefinite, while $P > 0$ and $0 < P$ indicate that P is positive definite. More generally, for Hermitian operators X and Y, one writes either $X \geq Y$ or $Y \leq X$ to indicate that $X - Y$ is positive semidefinite, and either $X > Y$ or $Y < X$ to indicate that $X - Y$ is positive definite.

Linear maps on square operators

Linear maps of the form

$$\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y}), \tag{1.105}$$

for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, play a fundamental role in the theory of quantum information. The set of all such maps is denoted $\mathrm{T}(\mathcal{X}, \mathcal{Y})$, and is itself a complex vector space when addition and scalar multiplication are defined in the straightforward way:

1. Addition: given two maps $\Phi, \Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, the map $\Phi + \Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is defined as
$$(\Phi + \Psi)(X) = \Phi(X) + \Psi(X) \tag{1.106}$$
for all $X \in \mathrm{L}(\mathcal{X})$.
2. Scalar multiplication: given a map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ and a scalar $\alpha \in \mathbb{C}$, the map $\alpha\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is defined as
$$(\alpha\Phi)(X) = \alpha\Phi(X) \tag{1.107}$$
for all $X \in \mathrm{L}(\mathcal{X})$.

For a given map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, the *adjoint* of Φ is defined to be the unique map $\Phi^* \in \mathrm{T}(\mathcal{Y}, \mathcal{X})$ that satisfies

$$\langle \Phi^*(Y), X \rangle = \langle Y, \Phi(X) \rangle \tag{1.108}$$

for all $X \in \mathrm{L}(\mathcal{X})$ and $Y \in \mathrm{L}(\mathcal{Y})$.

Tensor products of maps of the form (1.105) are defined in a similar way to tensor products of operators. More specifically, for any choice of complex Euclidean spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$ and $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ and linear maps

$$\Phi_1 \in \mathrm{T}(\mathcal{X}_1, \mathcal{Y}_1), \ \ldots, \ \Phi_n \in \mathrm{T}(\mathcal{X}_n, \mathcal{Y}_n), \tag{1.109}$$

one defines the tensor product of these maps

$$\Phi_1 \otimes \cdots \otimes \Phi_n \in \mathrm{T}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n) \tag{1.110}$$

to be the unique linear map that satisfies the equation

$$(\Phi_1 \otimes \cdots \otimes \Phi_n)(X_1 \otimes \cdots \otimes X_n) = \Phi_1(X_1) \otimes \cdots \otimes \Phi_n(X_n) \tag{1.111}$$

for all operators $X_1 \in \mathrm{L}(\mathcal{X}_1), \ldots, X_n \in \mathrm{L}(\mathcal{X}_n)$. As for vectors and operators, the notation $\Phi^{\otimes n}$ denotes the n-fold tensor product of a map Φ with itself.

The notation $\mathrm{T}(\mathcal{X})$ is a shorthand for $\mathrm{T}(\mathcal{X}, \mathcal{X})$. The identity map $\mathbb{1}_{\mathrm{L}(\mathcal{X})} \in \mathrm{T}(\mathcal{X})$ is defined as

$$\mathbb{1}_{\mathrm{L}(\mathcal{X})}(X) = X \tag{1.112}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

The trace function defined for square operators acting on $\mathcal{X}$ is a linear mapping of the form

$$\mathrm{Tr} : \mathrm{L}(\mathcal{X}) \to \mathbb{C}. \tag{1.113}$$

By making the identification $\mathrm{L}(\mathbb{C}) = \mathbb{C}$, one sees that the trace function is a linear map of the form

$$\mathrm{Tr} \in \mathrm{T}(\mathcal{X}, \mathbb{C}). \tag{1.114}$$

For a second complex Euclidean space $\mathcal{Y}$, one may consider the map

$$\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})} \in \mathrm{T}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Y}). \tag{1.115}$$

By the definition of the tensor product of maps stated above, this is the unique map that satisfies the equation

$$(\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(X \otimes Y) = \mathrm{Tr}(X) Y \tag{1.116}$$

for all operators $X \in \mathrm{L}(\mathcal{X})$ and $Y \in \mathrm{L}(\mathcal{Y})$. This map is called the *partial trace*, and is more commonly denoted $\mathrm{Tr}_{\mathcal{X}}$. Along similar lines, the map $\mathrm{Tr}_{\mathcal{Y}} \in \mathrm{T}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{X})$ is defined as

$$\mathrm{Tr}_{\mathcal{Y}} = \mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \mathrm{Tr}\,. \tag{1.117}$$

Generalizations of these maps may also be defined for tensor products of three or more complex Euclidean spaces.

The following classes of maps of the form (1.105) are among those that are discussed in greater detail later in this book.

1. *Hermiticity preserving maps.* A map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is *Hermiticity preserving* if and only if
$$\Phi(H) \in \mathrm{Herm}(\mathcal{Y}) \tag{1.118}$$
for every Hermitian operator $H \in \mathrm{Herm}(\mathcal{X})$.

2. *Positive maps.* A map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is *positive* if and only if
$$\Phi(P) \in \mathrm{Pos}(\mathcal{Y}) \tag{1.119}$$
for every positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X})$.

3. *Completely positive maps.* A map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is *completely positive* if and only if
$$\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})} \tag{1.120}$$
is a positive map for every complex Euclidean space $\mathcal{Z}$. The set of all completely positive maps of this form is denoted $\mathrm{CP}(\mathcal{X}, \mathcal{Y})$.

4. *Trace-preserving maps.* A map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is *trace-preserving* if and only if
$$\mathrm{Tr}(\Phi(X)) = \mathrm{Tr}(X) \tag{1.121}$$
for all $X \in \mathrm{L}(\mathcal{X})$.

5. *Unital maps.* A map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is *unital* if and only if
$$\Phi(\mathbb{1}_{\mathcal{X}}) = \mathbb{1}_{\mathcal{Y}}. \tag{1.122}$$

Maps of these sorts are discussed in greater detail in Chapters 2 and 4.

The operator-vector correspondence

There is a correspondence between the spaces $\mathrm{L}(\mathcal{Y}, \mathcal{X})$ and $\mathcal{X} \otimes \mathcal{Y}$, for any choice of complex Euclidean spaces $\mathcal{X} = \mathbb{C}^{\Sigma}$ and $\mathcal{Y} = \mathbb{C}^{\Gamma}$, that will be used repeatedly throughout this book. This correspondence is given by the linear mapping

$$\mathrm{vec} : \mathrm{L}(\mathcal{Y}, \mathcal{X}) \to \mathcal{X} \otimes \mathcal{Y}, \tag{1.123}$$

defined by the action

$$\mathrm{vec}(E_{a,b}) = e_a \otimes e_b \tag{1.124}$$

for all $a \in \Sigma$ and $b \in \Gamma$. In other words, this mapping is the change-of-basis taking the standard basis of $\mathrm{L}(\mathcal{Y}, \mathcal{X})$ to the standard basis of $\mathcal{X} \otimes \mathcal{Y}$. By linearity, it holds that

$$\mathrm{vec}(uv^*) = u \otimes \overline{v} \tag{1.125}$$

for $u \in \mathcal{X}$ and $v \in \mathcal{Y}$. This includes the special cases

$$\mathrm{vec}(u) = u \quad \text{and} \quad \mathrm{vec}(v^*) = \overline{v}, \tag{1.126}$$

obtained by setting $v = 1$ and $u = 1$, respectively.

The vec mapping is a linear bijection, which implies that every vector $u \in \mathcal{X} \otimes \mathcal{Y}$ uniquely determines an operator $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ that satisfies $\mathrm{vec}(A) = u$. It is also an isometry, in the sense that

$$\langle A, B \rangle = \langle \mathrm{vec}(A), \mathrm{vec}(B) \rangle \tag{1.127}$$

for all $A, B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$.

A few specific identities concerning the vec mapping will be especially useful throughout this book. One such identity is

$$(A_0 \otimes A_1)\,\mathrm{vec}(B) = \mathrm{vec}\big(A_0 B A_1^{\mathsf{T}}\big), \tag{1.128}$$

holding for all operators $A_0 \in \mathrm{L}(\mathcal{X}_0, \mathcal{Y}_0)$, $A_1 \in \mathrm{L}(\mathcal{X}_1, \mathcal{Y}_1)$, and $B \in \mathrm{L}(\mathcal{X}_1, \mathcal{X}_0)$, over all choices of complex Euclidean spaces $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$. Two more such identities are

$$\mathrm{Tr}_{\mathcal{Y}}\big(\mathrm{vec}(A)\,\mathrm{vec}(B)^*\big) = AB^*, \tag{1.129}$$

$$\mathrm{Tr}_{\mathcal{X}}\big(\mathrm{vec}(A)\,\mathrm{vec}(B)^*\big) = \big(B^* A\big)^{\mathsf{T}}, \tag{1.130}$$

which hold for all operators $A, B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$, over all choices of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$.

1.1.3 Operator decompositions and norms

Two operator decompositions—the *spectral decomposition* and *singular value decomposition*—along with various notions relating to these decompositions, are discussed in the present section. Among these related notions is a class of operator norms called *Schatten norms*, which include the trace norm, the Frobenius norm, and the spectral norm. These three norms will be used extensively throughout this book.

The spectral theorem

The *spectral theorem* establishes that every normal operator can be expressed as a linear combination of projections onto pairwise orthogonal subspaces. A formal statement of the spectral theorem follows.

Theorem 1.3 (Spectral theorem). *Let $X \in \mathrm{L}(\mathcal{X})$ be a normal operator, for $\mathcal{X}$ being a complex Euclidean space. There exists a positive integer m, nonzero projection operators $\Pi_1, \ldots, \Pi_m \in \mathrm{Proj}(\mathcal{X})$ satisfying $\Pi_1 + \cdots + \Pi_m = \mathbb{1}_{\mathcal{X}}$, and distinct complex numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ such that*

$$X = \sum_{k=1}^{m} \lambda_k \Pi_k. \tag{1.131}$$

Moreover, the projection operators $\Pi_1, \ldots, \Pi_m$ and scalars $\lambda_1, \ldots, \lambda_m$ are unique, up to their ordering: each value λ_k is an eigenvalue of X with multiplicity equal to the rank of Π_k, and Π_k is the projection operator onto the space spanned by the eigenvectors of X having corresponding eigenvalue λ_k.

The expression of a normal operator X in the form of the above equation (1.131) is called a *spectral decomposition* of X.

A simple corollary of the spectral theorem follows. It expresses essentially the same fact as the spectral theorem, but in a slightly different form that will sometimes be convenient to refer to later in the book.

Corollary 1.4. *Let $\mathcal{X}$ be a complex Euclidean space having dimension n, let $X \in \mathrm{L}(\mathcal{X})$ be a normal operator, and assume that $\mathrm{spec}(X) = \{\lambda_1, \ldots, \lambda_n\}$. There exists an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$ such that*

$$X = \sum_{k=1}^{n} \lambda_k x_k x_k^*. \tag{1.132}$$

It is evident from the expression (1.132), along with the requirement that the set $\{x_1, \ldots, x_n\}$ is an orthonormal basis, that each x_k is an eigenvector of X whose corresponding eigenvalue is λ_k. It is also evident that any operator X that is expressible in such a form as (1.132) is normal, implying that the condition of normality is equivalent to the existence of an orthonormal basis of eigenvectors.

On a few occasions later in the book, it will be convenient to index the eigenvectors and eigenvalues of a given normal operator $X \in \mathrm{L}(\mathbb{C}^\Sigma)$ by symbols in the alphabet Σ rather than by integers in the set $\{1, \ldots, n\}$ for $n = |\Sigma|$. It follows immediately from Corollary 1.4 that a normal operator $X \in \mathrm{L}(\mathbb{C}^\Sigma)$ may be expressed as

$$X = \sum_{a \in \Sigma} \lambda_a x_a x_a^* \tag{1.133}$$

for some choice of an orthonormal basis $\{x_a : a \in \Sigma\}$ of $\mathbb{C}^\Sigma$ and a collection of complex numbers $\{\lambda_a : a \in \Sigma\}$. Indeed, such an expression may be derived from (1.132) by associating symbols in the alphabet Σ with integers in the set $\{1, \ldots, n\}$ with respect to an arbitrarily chosen bijection.

It is convenient to refer to expressions of operators having either the forms (1.132) or (1.133) as *spectral decompositions*, despite the fact that they may differ slightly from the form (1.131). Unlike the form (1.131), the forms (1.132) and (1.133) are generally not unique. Along similar lines, the term *spectral theorem* is sometimes used to refer to the statement of Corollary 1.4, as opposed to the statement of Theorem 1.3. These conventions are followed throughout this book when there is no danger of any confusion resulting from their use.

An important theorem regarding spectral decompositions of commuting normal operators follows. It states that the same orthonormal basis of eigenvectors $\{x_1, \ldots, x_n\}$ may be chosen for any two normal operators under the assumption that they commute.

Theorem 1.5. *Let $\mathcal{X}$ be a complex Euclidean space having dimension n and let $X, Y \in \mathrm{L}(\mathcal{X})$ be normal operators for which $[X, Y] = 0$. There exists an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$ such that*

$$X = \sum_{k=1}^{n} \alpha_k x_k x_k^* \quad \text{and} \quad Y = \sum_{k=1}^{n} \beta_k x_k x_k^*, \tag{1.134}$$

for complex numbers $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ satisfying $\operatorname{spec}(X) = \{\alpha_1, \ldots, \alpha_n\}$ and $\operatorname{spec}(Y) = \{\beta_1, \ldots, \beta_n\}$.

Jordan–Hahn decompositions

Every Hermitian operator is normal and has real eigenvalues. It therefore follows from the spectral theorem (Theorem 1.3) that, for every Hermitian operator $H \in \mathrm{Herm}(\mathcal{X})$, there exists a positive integer m, nonzero projection operators $\Pi_1, \ldots, \Pi_m$ satisfying

$$\Pi_1 + \cdots + \Pi_m = \mathbb{1}_{\mathcal{X}}, \tag{1.135}$$

and real numbers $\lambda_1, \ldots, \lambda_m$ such that

$$H = \sum_{k=1}^{m} \lambda_k \Pi_k. \tag{1.136}$$

By defining operators

$$P = \sum_{k=1}^{n} \max\{\lambda_k, 0\}\, \Pi_k \quad \text{and} \quad Q = \sum_{k=1}^{n} \max\{-\lambda_k, 0\}\, \Pi_k\,, \tag{1.137}$$

one finds that

$$H = P - Q \tag{1.138}$$

for $P, Q \in \mathrm{Pos}(\mathcal{X})$ satisfying $PQ = 0$. The expression (1.138) of a given Hermitian operator H in this form, for positive semidefinite operators P and Q satisfying $PQ = 0$, is called a *Jordan–Hahn decomposition*. There is only one such expression for a given operator $H \in \mathrm{Herm}(\mathcal{X})$; the operators P and Q are uniquely defined by the requirements that $P, Q \in \mathrm{Pos}(\mathcal{X})$, $PQ = 0$, and $H = P - Q$.

Functions of normal operators

Every function of the form $f : \mathbb{C} \to \mathbb{C}$ may be extended to the set of normal operators in $\mathrm{L}(\mathcal{X})$, for a given complex Euclidean space $\mathcal{X}$, by means of the spectral theorem (Theorem 1.3). In particular, if $X \in \mathrm{L}(\mathcal{X})$ is normal and has the spectral decomposition (1.131), then one defines

$$f(X) = \sum_{k=1}^{m} f(\lambda_k)\Pi_k. \tag{1.139}$$

Naturally, functions defined only on subsets of scalars may be extended to normal operators whose eigenvalues are restricted accordingly.

The following examples of scalar functions extended to operators will be important later in this book.

1. For $r > 0$, the function $\lambda \mapsto \lambda^r$ is defined for every nonnegative real number $\lambda \in [0, \infty)$. For a positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X})$ having spectral decomposition

$$P = \sum_{k=1}^{m} \lambda_k \Pi_k, \tag{1.140}$$

for which it necessarily holds that $\lambda_k \geq 0$ for all $k \in \{1, \ldots, m\}$, one defines

$$P^r = \sum_{k=1}^{m} \lambda_k^r \, \Pi_k. \tag{1.141}$$

For positive integer values of r, it is evident that P^r coincides with the usual meaning of this expression given by operator multiplication.

The case that $r = 1/2$ is particularly common, and in this case one may write $\sqrt{P}$ to denote $P^{1/2}$. The operator $\sqrt{P}$ is the unique positive semidefinite operator that satisfies the equation

$$\sqrt{P}\sqrt{P} = P. \tag{1.142}$$

2. Along similar lines to the previous example, for any real number $r \in \mathbb{R}$, the function $\lambda \mapsto \lambda^r$ is defined for positive real values $\lambda \in (0, \infty)$. For a given positive definite operator $P \in \mathrm{Pd}(\mathcal{X})$, one defines P^r in a similar way to above.
3. The function $\lambda \mapsto \log(\lambda)$ is defined for every positive real number $\lambda \in (0, \infty)$. For a given positive definite operator $P \in \mathrm{Pd}(\mathcal{X})$, having a spectral decomposition (1.140) as above, one defines

$$\log(P) = \sum_{k=1}^{m} \log(\lambda_k) \Pi_k. \tag{1.143}$$

The singular-value theorem

The *singular value theorem* has a close relationship to the spectral theorem. Unlike the spectral theorem, the singular value theorem holds for arbitrary (nonzero) operators, as opposed to just normal operators.

Theorem 1.6 (Singular value theorem). *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ be a nonzero operator having rank equal to r. There exist positive real numbers $s_1, \ldots, s_r$ and orthonormal sets $\{x_1, \ldots, x_r\} \subset \mathcal{X}$ and $\{y_1, \ldots, y_r\} \subset \mathcal{Y}$ such that*

$$A = \sum_{k=1}^{r} s_k y_k x_k^*. \tag{1.144}$$

An expression of a given operator A in the form of (1.144) is said to be a *singular value decomposition* of A. The numbers $s_1, \ldots, s_r$ are called *singular values* and the vectors $x_1, \ldots, x_r$ and $y_1, \ldots, y_r$ are called *right* and *left singular vectors*, respectively.

The singular values $s_1, \ldots, s_r$ of an operator A are uniquely determined, up to their ordering. It will be assumed hereafter that singular values are always ordered from largest to smallest: $s_1 \geq \cdots \geq s_r$. When it is necessary to indicate the dependence of these singular values on the operator A, they are denoted $s_1(A), \ldots, s_r(A)$. Although 0 is not formally considered to be a singular value of any operator, it is convenient to also define $s_k(A) = 0$ for $k > \operatorname{rank}(A)$, and to take $s_k(A) = 0$ for all $k \geq 1$ when $A = 0$. The notation $s(A)$ is used to refer to the vector of singular values

$$s(A) = (s_1(A), \ldots, s_r(A)), \tag{1.145}$$

or to an extension of this vector

$$s(A) = (s_1(A), \ldots, s_m(A)) \tag{1.146}$$

for $m > r$, when it is convenient to view this vector as an element of $\mathbb{R}^m$ for $m > \operatorname{rank}(A)$.

As suggested above, there is a close relationship between the singular value theorem and the spectral theorem. In particular, the singular value decomposition of an operator A and the spectral decompositions of the operators A^*A and AA^* are related in the following way: it necessarily holds that

$$s_k(A) = \sqrt{\lambda_k(AA^*)} = \sqrt{\lambda_k(A^*A)} \tag{1.147}$$

for $1 \leq k \leq \operatorname{rank}(A)$, and moreover the right singular vectors of A are eigenvectors of A^*A and the left singular vectors of A are eigenvectors of AA^*. One is free, in fact, to choose the left singular vectors of A to be any orthonormal collection of eigenvectors of AA^* for which the corresponding

eigenvalues are nonzero—and once this is done the right singular vectors will be uniquely determined. Alternately, the right singular vectors of A may be chosen to be any orthonormal collection of eigenvectors of A^*A for which the corresponding eigenvalues are nonzero, which uniquely determines the left singular vectors.

In the special case that $X \in \mathrm{L}(\mathcal{X})$ is a normal operator, one may obtain a singular value decomposition of X directly from a spectral decomposition of the form

$$X = \sum_{k=1}^{n} \lambda_k x_k x_k^*. \tag{1.148}$$

In particular, one may define $S = \{k \in \{1, \ldots, n\} : \lambda_k \neq 0\}$, and set

$$s_k = |\lambda_k| \quad \text{and} \quad y_k = \frac{\lambda_k}{|\lambda_k|} x_k \tag{1.149}$$

for each $k \in S$. The expression

$$X = \sum_{k \in S} s_k y_k x_k^* \tag{1.150}$$

then represents a singular value decomposition of X, up to a relabeling of the terms in the sum.

The following corollary represents a reformulation of the singular value theorem that is useful in some situations.

Corollary 1.7. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ be a nonzero operator, and let $r = \operatorname{rank}(A)$. There exists a positive definite, diagonal operator $D \in \mathrm{Pd}(\mathbb{C}^r)$ and linear isometries $U \in \mathrm{U}(\mathbb{C}^r, \mathcal{X})$ and $V \in \mathrm{U}(\mathbb{C}^r, \mathcal{Y})$ such that $A = VDU^*$.*

Polar decompositions

For every square operator $X \in \mathrm{L}(\mathcal{X})$, it is possible to choose a positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X})$ and a unitary operator $W \in \mathrm{U}(\mathcal{X})$ such that the equation

$$X = WP \tag{1.151}$$

holds; this follows from Corollary 1.7 by taking $W = VU^*$ and $P = UDU^*$. Alternatively, by similar reasoning it is possible to write

$$X = PW \tag{1.152}$$

for a (generally different) choice of operators $P \in \mathrm{Pos}(\mathcal{X})$ and $W \in \mathrm{U}(\mathcal{X})$. The expressions (1.151) and (1.152) are known as *polar decompositions* of X.

Schmidt decompositions

Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and suppose that $u \in \mathcal{X} \otimes \mathcal{Y}$ is a nonzero vector. Given that the vec mapping is a bijection, there exists a unique operator $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ such that $u = \mathrm{vec}(A)$. For any singular value decomposition

$$A = \sum_{k=1}^{r} s_k x_k y_k^*, \tag{1.153}$$

it holds that

$$u = \mathrm{vec}(A) = \mathrm{vec}\left(\sum_{k=1}^{r} s_k x_k y_k^*\right) = \sum_{k=1}^{r} s_k x_k \otimes \overline{y_k}. \tag{1.154}$$

The orthonormality of $\{y_1, \ldots, y_r\}$ implies that $\{\overline{y_1}, \ldots, \overline{y_r}\}$ is orthonormal as well. It follows that every nonzero vector $u \in \mathcal{X} \otimes \mathcal{Y}$ can be expressed in the form

$$u = \sum_{k=1}^{r} s_k x_k \otimes z_k \tag{1.155}$$

for positive real numbers $s_1, \ldots, s_r$ and orthonormal sets $\{x_1, \ldots, x_r\} \subset \mathcal{X}$ and $\{z_1, \ldots, z_r\} \subset \mathcal{Y}$. An expression of u having this form is called a *Schmidt decomposition* of u.

The Moore–Penrose pseudo-inverse

For a given operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, one defines an operator $A^+ \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$, known as the *Moore–Penrose pseudo-inverse* of A, as the unique operator that possesses the following properties:

1. $AA^+A = A$,
2. $A^+AA^+ = A^+$, and
3. AA^+ and A^+A are both Hermitian.

It is evident that there is at least one such choice of A^+, for if

$$A = \sum_{k=1}^{r} s_k y_k x_k^* \tag{1.156}$$

is a singular value decomposition of a nonzero operator A, then

$$A^+ = \sum_{k=1}^{r} \frac{1}{s_k} x_k y_k^* \tag{1.157}$$

possesses the three properties listed above. One may observe that AA^+ and A^+A are projection operators, projecting onto the spaces spanned by the left singular vectors and right singular vectors of A, respectively.

The fact that A^+ is uniquely determined by the above equations may be verified as follows. Suppose that $B, C \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ both possess the above properties:

1. $ABA = A = ACA$,
2. $BAB = B$ and $CAC = C$, and
3. AB, BA, AC, and CA are all Hermitian.

It follows that

$$\begin{aligned} B &= BAB = (BA)^*B = A^*B^*B = (ACA)^*B^*B \\ &= A^*C^*A^*B^*B = (CA)^*(BA)^*B = CABAB \\ &= CAB = CACAB = C(AC)^*(AB)^* = CC^*A^*B^*A^* \\ &= CC^*(ABA)^* = CC^*A^* = C(AC)^* = CAC = C, \end{aligned} \tag{1.158}$$

which shows that $B = C$.

Norms of operators

A *norm* on the space of operators $\mathrm{L}(\mathcal{X}, \mathcal{Y})$, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, is a function $\|\cdot\|$ satisfying the following properties:

1. Positive definiteness: $\|A\| \geq 0$ for all $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, with $\|A\| = 0$ if and only if $A = 0$.
2. Positive scalability: $\|\alpha A\| = |\alpha| \|A\|$ for all $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and $\alpha \in \mathbb{C}$.
3. The triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$ for all $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$.

Many interesting and useful norms can be defined on spaces of operators, but this book will mostly be concerned with a single family of norms called *Schatten p-norms*. This family includes the three most commonly used norms in quantum information theory: the *spectral norm*, the *Frobenius norm*, and the *trace norm*.

For any operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ and any real number $p \geq 1$, one defines the Schatten p-norm of A as

$$\|A\|_p = \left(\mathrm{Tr}\left((A^*A)^{\frac{p}{2}}\right)\right)^{\frac{1}{p}}. \tag{1.159}$$

The Schatten ∞-norm is defined as

$$\|A\|_\infty = \max\{\|Au\| \,:\, u \in \mathcal{X},\ \|u\| = 1\}, \tag{1.160}$$

which coincides with $\lim_{p\to\infty}\|A\|_p$, explaining why the subscript ∞ is used. The Schatten p-norm of an operator A coincides with the ordinary vector p-norm of the vector of singular values of A:

$$\|A\|_p = \|s(A)\|_p. \tag{1.161}$$

The Schatten p-norms possess a variety of properties, including the ones summarized in the following list:

1. The Schatten p-norms are non-increasing in p: for every $A \in \mathrm{L}(\mathcal{X},\mathcal{Y})$ and for $1 \leq p \leq q \leq \infty$, it holds that

   $$\|A\|_p \geq \|A\|_q. \tag{1.162}$$

   For every nonzero operator A it holds that

   $$\|A\|_p \leq \mathrm{rank}(A)^{\frac{1}{p}-\frac{1}{q}}\|A\|_q. \tag{1.163}$$

   In particular, one has

   $$\|A\|_1 \leq \sqrt{\mathrm{rank}(A)}\,\|A\|_2 \quad \text{and} \quad \|A\|_2 \leq \sqrt{\mathrm{rank}(A)}\,\|A\|_\infty. \tag{1.164}$$

2. For every $p \in [1,\infty]$, the Schatten p-norm is isometrically invariant (and therefore unitarily invariant): for every $A \in \mathrm{L}(\mathcal{X},\mathcal{Y})$, $U \in \mathrm{U}(\mathcal{Y},\mathcal{Z})$, and $V \in \mathrm{U}(\mathcal{X},\mathcal{W})$ it holds that

   $$\|A\|_p = \|UAV^*\|_p. \tag{1.165}$$

3. For each $p \in [1,\infty]$, one defines $p^* \in [1,\infty]$ by the equation

   $$\frac{1}{p} + \frac{1}{p^*} = 1. \tag{1.166}$$

   For every operator $A \in \mathrm{L}(\mathcal{X},\mathcal{Y})$, it holds that the Schatten p-norm and p^*-norm are dual, in the sense that

   $$\|A\|_p = \max\bigl\{|\langle B, A\rangle| \,:\, B \in \mathrm{L}(\mathcal{X},\mathcal{Y}),\ \|B\|_{p^*} \leq 1\bigr\}. \tag{1.167}$$

One consequence of (1.167) is the inequality

$$|\langle B, A\rangle| \leq \|A\|_p \|B\|_{p^*}, \tag{1.168}$$

which is known as the *Hölder inequality* for Schatten norms.

4. For every choice of operators $A \in \mathrm{L}(\mathcal{Z}, \mathcal{W})$, $B \in \mathrm{L}(\mathcal{Y}, \mathcal{Z})$, and $C \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, and any choice of $p \in [1, \infty]$, it holds that

$$\|ABC\|_p \leq \|A\|_\infty \|B\|_p \|C\|_\infty. \tag{1.169}$$

It follows that the Schatten p-norm is *submultiplicative*:

$$\|AB\|_p \leq \|A\|_p \|B\|_p. \tag{1.170}$$

5. For every $p \in [0, \infty]$ and every $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$, it holds that

$$\|A\|_p = \left\|A^*\right\|_p = \left\|A^{\mathsf{T}}\right\|_p = \left\|\overline{A}\right\|_p. \tag{1.171}$$

The Schatten 1-norm is commonly called the *trace norm*, the Schatten 2-norm is also known as the *Frobenius norm*, and the Schatten ∞-norm is called the *spectral norm* or *operator norm*. Some aditional properties of these three norms are as follows.

1. *The spectral norm.* The spectral norm $\|\cdot\|_\infty$ is special in several respects. It is the norm *induced* by the Euclidean norm, which is its defining property (1.160). It also satisfies the property

$$\|A^*A\|_\infty = \|AA^*\|_\infty = \|A\|_\infty^2 \tag{1.172}$$

for every $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$. Hereafter in this book, the spectral norm of an operator A will be written $\|A\|$ rather than $\|A\|_\infty$, which reflects the fundamental nature of this norm.

2. *The Frobenius norm.* Substituting $p = 2$ into the definition of $\|\cdot\|_p$, one sees that the Frobenius norm $\|\cdot\|_2$ is given by

$$\|A\|_2 = \big(\mathrm{Tr}(A^*A)\big)^{\frac{1}{2}} = \sqrt{\langle A, A\rangle}, \tag{1.173}$$

and is therefore analogous to the Euclidean norm for vectors, but defined by the inner product on $\mathrm{L}(\mathcal{X}, \mathcal{Y})$.

In essence, the Frobenius norm corresponds to the Euclidean norm of an operator viewed as a vector:

$$\|A\|_2 = \|\mathrm{vec}(A)\| = \sqrt{\sum_{a,b} |A(a,b)|^2}, \tag{1.174}$$

where a and b range over the indices of the matrix representation of A.

3. *The trace norm.* Substituting $p = 1$ into the definition of $\|\cdot\|_p$, one has that the trace norm $\|\cdot\|_1$ is given by

$$\|A\|_1 = \mathrm{Tr}\Big(\sqrt{A^*A}\Big), \tag{1.175}$$

which is equal to the sum of the singular values of A. A useful expression of $\|X\|_1$, for any square operator $X \in \mathrm{L}(\mathcal{X})$, is

$$\|X\|_1 = \max\big\{|\langle U, X\rangle| \,:\, U \in \mathrm{U}(\mathcal{X})\big\}, \tag{1.176}$$

which follows from (1.161) together with the singular value theorem (Theorem 1.6). As a result, one has that the trace-norm is non-increasing under the action of partial tracing: for every operator $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$, it holds that

$$\begin{aligned} \|\mathrm{Tr}_{\mathcal{Y}}(X)\|_1 &= \max\big\{\langle U \otimes \mathbb{1}_{\mathcal{Y}}, X\rangle \,:\, U \in \mathrm{U}(\mathcal{X})\big\} \\ &\leq \max\big\{\langle V, X\rangle \,:\, V \in \mathrm{U}(\mathcal{X} \otimes \mathcal{Y})\big\} = \|X\|_1. \end{aligned} \tag{1.177}$$

The identity

$$\big\|\alpha uu^* - \beta vv^*\big\|_1 = \sqrt{(\alpha+\beta)^2 - 4\alpha\beta|\langle u, v\rangle|^2}, \tag{1.178}$$

which holds for any choice of unit vectors u, v and nonnegative real numbers α, β, is used multiple times in this book. It may be proved by considering the spectrum of $\alpha uu^* - \beta vv^*$; this operator is Hermitian, and has at most two nonzero eigenvalues, represented by the expression

$$\frac{\alpha - \beta}{2} \pm \frac{1}{2}\sqrt{(\alpha+\beta)^2 - 4\alpha\beta|\langle u, v\rangle|^2}. \tag{1.179}$$

In particular, for unit vectors u and v, one has

$$\big\|uu^* - vv^*\big\|_1 = 2\sqrt{1 - |\langle u, v\rangle|^2}. \tag{1.180}$$

1.2 Analysis, convexity, and probability theory

Some of the proofs to be presented in this book will make use of concepts from analysis, convexity, and probability theory. The summary that follows provides an overview of these concepts, narrowly focused on the needs of this book.

1.2.1 Analysis and convexity

In the same spirit as the previous section on linear algebra, it is assumed that the reader is familiar with the most basic notions of mathematical analysis, including the supremum and infimum of sets of real numbers, sequences and limits, and standard univariate calculus over the real numbers.

The discussion below is limited to finite-dimensional real and complex vector spaces—and the reader is cautioned that some of the stated facts rely on the assumption that one is working with finite dimensional spaces. For the remainder of the subsection, $\mathcal{V}$ and $\mathcal{W}$ will denote finite dimensional real or complex vector spaces, upon which some particular norm (which may be chosen arbitrarily) has been fixed. These norms are denoted by the usual norm symbol $\|\cdot\|$, which therefore does not necessarily refer to the Euclidean norm or spectral norm in this context.

Open and closed sets

A set $\mathcal{A} \subseteq \mathcal{V}$ is *open* if and only if, for every $u \in \mathcal{A}$, there exists a choice of $\varepsilon > 0$ such that

$$\big\{v \in \mathcal{V} \,:\, \|u - v\| < \varepsilon\big\} \subseteq \mathcal{A}. \tag{1.181}$$

A set $\mathcal{A} \subseteq \mathcal{V}$ is *closed* if and only if the complement of $\mathcal{A}$, defined as

$$\mathcal{V} \backslash \mathcal{A} = \big\{v \in \mathcal{V} \,:\, v \notin \mathcal{A}\big\}, \tag{1.182}$$

is open. Given subsets $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$, one defines that $\mathcal{A}$ is open or closed *relative to* $\mathcal{B}$ if $\mathcal{A}$ is the intersection of $\mathcal{B}$ with some set in $\mathcal{V}$ that is open or closed, respectively. Equivalently, $\mathcal{A}$ is open relative to $\mathcal{B}$ if and only if, for every $u \in \mathcal{A}$, there exists a choice of $\varepsilon > 0$ such that

$$\big\{v \in \mathcal{B} \,:\, \|u - v\| < \varepsilon\big\} \subseteq \mathcal{A}; \tag{1.183}$$

and $\mathcal{A}$ is closed relative to $\mathcal{B}$ if and only if $\mathcal{B} \backslash \mathcal{A}$ is open relative to $\mathcal{B}$.

For subsets $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$, one defines the *closure* of $\mathcal{A}$ relative to $\mathcal{B}$ as the intersection of all subsets $\mathcal{C}$ such that $\mathcal{A} \subseteq \mathcal{C} \subseteq \mathcal{B}$ and $\mathcal{C}$ is closed relative to $\mathcal{B}$. In other words, this is the smallest set that contains $\mathcal{A}$ and is closed relative to $\mathcal{B}$. The set $\mathcal{A}$ is *dense* in $\mathcal{B}$ if the closure of $\mathcal{A}$ relative to $\mathcal{B}$ is $\mathcal{B}$ itself.

Continuous functions

Let $f : \mathcal{A} \to \mathcal{W}$ be a function defined on some subset $\mathcal{A} \subseteq \mathcal{V}$. For any point $u \in \mathcal{A}$, the function f is said to be *continuous* at u if the following holds: for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|f(v) - f(u)\| < \varepsilon \tag{1.184}$$

for all $v \in \mathcal{A}$ satisfying $\|u - v\| < \delta$. If f is continuous at every point in $\mathcal{A}$, then one simply says that f is *continuous on* $\mathcal{A}$.

For a function $f : \mathcal{A} \to \mathcal{W}$ defined on some subset $\mathcal{A} \subseteq \mathcal{V}$, the *preimage* of a set $\mathcal{B} \subseteq \mathcal{W}$ is defined as

$$f^{-1}(\mathcal{B}) = \big\{u \in \mathcal{A} \,:\, f(u) \in \mathcal{B}\big\}. \tag{1.185}$$

Such a function f is continuous on $\mathcal{A}$ if and only if the preimage of every open set in $\mathcal{W}$ is open relative to $\mathcal{A}$. Equivalently, f is continuous on $\mathcal{A}$ if and only if the preimage of every closed set in $\mathcal{W}$ is closed relative to $\mathcal{A}$.

For a positive real number κ, a function $f : \mathcal{A} \to \mathcal{W}$ defined on a subset $A \subseteq \mathcal{V}$ is said to be a *κ-Lipschitz function* if and only if

$$\|f(u) - f(v)\| \leq \kappa \|u - v\| \tag{1.186}$$

for all $u, v \in \mathcal{A}$. Every κ-Lipschitz function is necessarily continuous.

Compact sets

A set $\mathcal{A} \subseteq \mathcal{V}$ is *compact* if and only if every sequence in $\mathcal{A}$ has a subsequence that converges to a vector $u \in \mathcal{V}$. As a consequence of the fact $\mathcal{V}$ is assumed to be finite dimensional, one has that a set $\mathcal{A} \subseteq \mathcal{V}$ is compact if and only if it is both closed and bounded—a fact known as the *Heine–Borel theorem*.

Two properties regarding continuous functions and compact sets that are particularly noteworthy for the purposes of this book are as follows.

1. If $\mathcal{A}$ is compact and $f : \mathcal{A} \to \mathbb{R}$ is continuous on $\mathcal{A}$, then f achieves both a maximum and minimum value on $\mathcal{A}$.

2. If $\mathcal{A} \subset \mathcal{V}$ is compact and $f : \mathcal{V} \to \mathcal{W}$ is continuous on $\mathcal{A}$, then

$$f(\mathcal{A}) = \{f(u) \,:\, u \in \mathcal{A}\} \tag{1.187}$$

is also compact. In words, continuous functions always map compact sets to compact sets.

Differentiation of multivariate real functions

Basic multivariate calculus will be employed in a few occasions later in this book, and in these cases it will be sufficient to consider only real-valued functions.

Suppose n is a positive integer, $f : \mathbb{R}^n \to \mathbb{R}$ is a function, and $x \in \mathbb{R}^n$ is a vector. Under the assumption that the partial derivative

$$\partial_k f(x) = \lim_{\alpha \to 0} \frac{f(x + \alpha e_k) - f(x)}{\alpha} \tag{1.188}$$

exists and is finite for each $k \in \{1, \ldots, n\}$, one defines the *gradient vector* of f at x as

$$\nabla f(x) = \big(\partial_1 f(x), \ldots, \partial_n f(x)\big). \tag{1.189}$$

If it holds that f is a κ-Lipschitz function and the gradient $\nabla f(x)$ of f at x is defined, then it must hold that

$$\|\nabla f(x)\| \leq \kappa. \tag{1.190}$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *differentiable* at a vector $x \in \mathbb{R}^n$ if and only if there exists a vector $y \in \mathbb{R}^n$ with the following property: for every sequence $(v_1, v_2, \ldots)$ of vectors in $\mathbb{R}^n$ that converges to 0, one has that

$$\lim_{k \to \infty} \frac{|f(x + v_k) - f(x) - \langle y, v_k \rangle|}{\|v_k\|} = 0. \tag{1.191}$$

In this case the vector y is necessarily unique, and one writes $y = (Df)(x)$. If f is differentiable at x, then it holds that

$$(Df)(x) = \nabla f(x). \tag{1.192}$$

It may be the case that the gradient vector $\nabla f(x)$ is defined for a vector x at which f is not differentiable, but if the function $x \mapsto \nabla f(x)$ is continuous at x, then f is necessarily differentiable at x.

Finally, suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function that is differentiable at a vector $x \in \mathbb{R}^n$, $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ are functions that are differentiable at a real number $\alpha \in \mathbb{R}$, and assume that

$$x = (g_1(\alpha), \ldots, g_n(\alpha)). \tag{1.193}$$

The *chain rule* for differentiation implies that the function $h : \mathbb{R} \to \mathbb{R}$ defined as

$$h(\beta) = f(g_1(\beta), \ldots, g_n(\beta)) \tag{1.194}$$

is differentiable at α, with derivative given by

$$h'(\alpha) = \big\langle \nabla f(x), (g_1'(\alpha), \ldots, g_n'(\alpha)) \big\rangle \tag{1.195}$$

Nets

Let $\mathcal{V}$ be a real or complex vector space, let $\mathcal{A} \subseteq \mathcal{V}$ be a subset of $\mathcal{V}$, let $\|\cdot\|$ be a norm on $\mathcal{V}$, and let $\varepsilon > 0$ be a positive real number. A set of vectors $\mathcal{N} \subseteq \mathcal{V}$ is an *ε-net* for $\mathcal{A}$ if and only if, for every vector $u \in \mathcal{A}$, there exists a vector $v \in \mathcal{N}$ such that $\|u - v\| \leq \varepsilon$. An ε-net $\mathcal{N}$ for $\mathcal{A}$ is *minimal* if and only if $\mathcal{N}$ is finite, and every ε-net of $\mathcal{A}$ contains at least $|\mathcal{N}|$ vectors.

The following theorem gives an upper bound for the number of elements in a minimal ε-net for the unit ball

$$\mathcal{B}(\mathcal{X}) = \{x \in \mathcal{X} : \|x\| \leq 1\} \tag{1.196}$$

in a complex Euclidean space, with respect to the Euclidean norm.

Theorem 1.8 (Pisier)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let $n = \dim(\mathcal{X})$, and let $\varepsilon > 0$ be a positive real number. With respect to the Euclidean norm on $\mathcal{X}$, there exists an ε-net $\mathcal{N} \subset \mathcal{B}(\mathcal{X})$ for the unit ball $\mathcal{B}(\mathcal{X})$ such that*

$$|\mathcal{N}| \leq \left(1 + \frac{2}{\varepsilon}\right)^{2n}. \tag{1.197}$$

The proof of this theorem does not require a complicated construction; one may take $\mathcal{N}$ to be any maximal set of vectors chosen from the unit ball for which it holds that $\|u - v\| \geq \varepsilon$ for all $u, v \in \mathcal{N}$ with $u \neq v$. Such a set is necessarily an ε-net for $\mathcal{B}(\mathcal{X})$, and the bound on $|\mathcal{N}|$ is obtained by comparing the volume of $\mathcal{B}(\mathcal{X})$ with the volume of the union of $\varepsilon/2$ balls around vectors in $\mathcal{N}$ (which are disjoint sets).

Borel sets and functions

Throughout this subsection, $\mathcal{A} \subseteq \mathcal{V}$ and $\mathcal{B} \subseteq \mathcal{W}$ will denote fixed subsets of (not necessarily equal) finite-dimensional real or complex vector spaces $\mathcal{V}$ and $\mathcal{W}$.

A set $\mathcal{C} \subseteq \mathcal{A}$ is said to be a *Borel subset* of $\mathcal{A}$ if one or more of the following inductively defined properties holds:

1. $\mathcal{C}$ is an open set relative to $\mathcal{A}$.
2. $\mathcal{C}$ is the complement of a Borel subset of $\mathcal{A}$.
3. $\mathcal{C}$ is equal to the union of a countable collection of Borel subsets of $\mathcal{A}$:

$$\mathcal{C} = \bigcup_{k=1}^{\infty} \mathcal{C}_k \tag{1.198}$$

for $\{\mathcal{C}_1, \mathcal{C}_2, \ldots\}$ being any countable collection of Borel subsets of $\mathcal{A}$.

The collection of all Borel subsets of $\mathcal{A}$ is denoted $\text{Borel}(\mathcal{A})$.

Generally speaking, the Borel subsets of $\mathcal{A}$ represent a rich collection of sets. For instance, any open or closed set relative to $\mathcal{A}$ is a Borel set, as is any set that can be obtained by taking countable unions, countable intersections, and complements of open and closed sets relative to $\mathcal{A}$.

A function $f : \mathcal{A} \to \mathcal{B}$ is a *Borel function* if and only if the preimage of every Borel subset of $\mathcal{B}$ is a Borel subset of $\mathcal{A}$, i.e., $f^{-1}(\mathcal{C}) \in \text{Borel}(\mathcal{A})$ for all $\mathcal{C} \in \text{Borel}(\mathcal{B})$. If f is a continuous function, then f is necessarily a Borel function as well. Another important type of Borel function is any function of the form

$$f(x) = \chi_{\mathcal{C}}(x)\, y \tag{1.199}$$

for any choice of $y \in \mathcal{B}$ and

$$\chi_{\mathcal{C}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{C} \\ 0 & \text{if } x \notin \mathcal{C} \end{cases} \tag{1.200}$$

being the characteristic function of any Borel subset $\mathcal{C} \in \text{Borel}(\mathcal{A})$.

The collection of all Borel functions $f : \mathcal{A} \to \mathcal{B}$ possesses a variety of closure properties, including the following properties:

1. If $\mathcal{B}$ is a vector space, $f, g : \mathcal{A} \to \mathcal{B}$ are Borel functions, and α is a scalar (either real or complex, depending on whether $\mathcal{B}$ is a real or complex vector space), then the functions αf and $f + g$ are also Borel functions.

2. If $\mathcal{B}$ is a subalgebra of $\mathrm{L}(\mathcal{Z})$, for $\mathcal{Z}$ being a real or complex Euclidean space, and $f, g : \mathcal{A} \to \mathcal{B}$ are Borel functions, then the function $h : \mathcal{A} \to \mathcal{B}$ defined by $h(x) = f(x)g(x)$ for all $x \in \mathcal{A}$ is also a Borel function. (This includes the special cases $f, g : \mathcal{A} \to \mathbb{R}$ and $f, g : \mathcal{A} \to \mathbb{C}$.)

Measures on Borel sets

A *Borel measure* (or simply a *measure*) defined on $\mathrm{Borel}(\mathcal{A})$ is a function

$$\mu : \mathrm{Borel}(\mathcal{A}) \to [0, \infty] \tag{1.201}$$

that possesses two properties:

1. $\mu(\varnothing) = 0$.
2. For any countable collection $\{\mathcal{C}_1, \mathcal{C}_2, \ldots\} \subseteq \mathrm{Borel}(\mathcal{A})$ of pairwise disjoint Borel subsets of $\mathcal{A}$, it holds that

$$\mu\left(\bigcup_{k=1}^{\infty} \mathcal{C}_k\right) = \sum_{k=1}^{\infty} \mu(\mathcal{C}_k). \tag{1.202}$$

A measure μ defined on $\mathrm{Borel}(\mathcal{A})$ is said to be *normalized* if it holds that $\mu(\mathcal{A}) = 1$. The term *probability measure* is also used to refer to a normalized measure.

There exists a measure ν defined on $\mathrm{Borel}(\mathbb{R})$, known as the *standard Borel measure*,[3] that has the property

$$\nu([\alpha, \beta]) = \beta - \alpha \tag{1.203}$$

for all choices of $\alpha, \beta \in \mathbb{R}$ with $\alpha \leq \beta$.

If $\mathcal{A}_1, \ldots, \mathcal{A}_n$ are subsets of (not necessarily equal) finite-dimensional real or complex vector spaces, and

$$\mu_k : \mathrm{Borel}(\mathcal{A}_k) \to [0, \infty] \tag{1.204}$$

is a measure for each $k \in \{1, \ldots, n\}$, then there is a uniquely defined *product measure*

$$\mu_1 \times \cdots \times \mu_n : \mathrm{Borel}(\mathcal{A}_1 \times \cdots \times \mathcal{A}_n) \to [0, \infty] \tag{1.205}$$

[3] The standard Borel measure agrees with the well-known *Lebesgue measure* on every Borel subset of $\mathbb{R}$. The Lebesgue measure is also defined for some subsets of $\mathbb{R}$ that are not Borel subsets, which endows it with additional properties that are not relevant within the context of this book.

for which

$$(\mu_1 \times \cdots \times \mu_n)(\mathcal{B}_1 \times \cdots \times \mathcal{B}_n) = \mu_1(\mathcal{B}_1) \cdots \mu_n(\mathcal{B}_n) \tag{1.206}$$

for all $\mathcal{B}_1 \in \text{Borel}(\mathcal{A}_1), \ldots, \mathcal{B}_n \in \text{Borel}(\mathcal{A}_n)$.

Integration of Borel functions

For some (but not all) Borel functions $f : \mathcal{A} \to \mathcal{B}$, and for μ being a Borel measure of the form $\mu : \text{Borel}(\mathcal{A}) \to [0, \infty]$, one may define the integral

$$\int f(x) \, \mathrm{d}\mu(x), \tag{1.207}$$

which is an element of $\mathcal{B}$ when it is defined.

An understanding of the specifics of the definition through which such an integral is defined is not critical within the context of this book, but some readers may find that a high-level overview of the definition is helpful in associating an intuitive meaning to the integrals that do arise. In short, one defines what is meant by the integral of an increasingly large collection of functions, beginning with functions taking nonnegative real values, and then proceeding to vector (or operator) valued functions by taking linear combinations.

1. *Nonnegative simple functions.* A function $g : \mathcal{A} \to [0, \infty)$ is a *nonnegative simple function* if and only if it may be written as

   $$g(x) = \sum_{k=1}^{m} \alpha_k \chi_k(x) \tag{1.208}$$

   for a nonnegative integer m, distinct positive real numbers $\alpha_1, \ldots, \alpha_m$, and characteristic functions $\chi_1, \ldots, \chi_k$ given by

   $$\chi_k(x) = \begin{cases} 1 & \text{if } x \in \mathcal{C}_k \\ 0 & \text{if } x \notin \mathcal{C}_k \end{cases} \tag{1.209}$$

   for disjoint Borel sets $\mathcal{C}_1, \ldots, \mathcal{C}_m \in \text{Borel}(\mathcal{A})$. (It is to be understood that the sum is empty when $m = 0$, which corresponds to g being identically zero.)

   A nonnegative simple function g of the form (1.208) is *integrable* with respect to a measure $\mu : \text{Borel}(\mathcal{A}) \to [0, \infty]$ if and only if $\mu(\mathcal{C}_k)$ is finite

for every $k \in \{1, \ldots, m\}$, and in this case the integral of g with respect to μ is defined as

$$\int g(x) \, \mathrm{d}\mu(x) = \sum_{k=1}^{m} \alpha_k \, \mu(\mathcal{C}_k). \tag{1.210}$$

This is a well-defined quantity, by virtue of the fact that the expression (1.208) happens to be unique for a given simple function g.

2. *Nonnegative Borel functions.* The integral of a Borel function of the form $f : \mathcal{A} \to [0, \infty)$, with respect to a given measure $\mu : \mathrm{Borel}(\mathcal{A}) \to [0, \infty]$, is defined as

$$\int f(x) \, \mathrm{d}\mu(x) = \sup \int g(x) \, \mathrm{d}\mu(x), \tag{1.211}$$

where the supremum is taken over all nonnegative simple functions of the form $g : \mathcal{A} \to [0, \infty)$ for which it holds that $g(x) \leq f(x)$ for all $x \in \mathcal{A}$. It is said that f is *integrable* if and only if the supremum value in (1.211) is finite.

3. *Real and complex Borel functions.* A Borel function $g : \mathcal{A} \to \mathbb{R}$ is *integrable* with respect to a measure $\mu : \mathrm{Borel}(\mathcal{A}) \to [0, \infty]$ if and only if there exist integrable Borel functions $f_0, f_1 : \mathcal{A} \to [0, \infty)$ such that $g = f_0 - f_1$, and in this case the integral of g with respect to μ is defined as

$$\int g(x) \, \mathrm{d}\mu(x) = \int f_0(x) \, \mathrm{d}\mu(x) - \int f_1(x) \, \mathrm{d}\mu(x). \tag{1.212}$$

Along similar lines, a Borel function $h : \mathcal{A} \to \mathbb{C}$ is *integrable* with respect to a measure $\mu : \mathrm{Borel}(\mathcal{A}) \to [0, \infty]$ if and only if there exist integrable Borel functions $g_0, g_1 : \mathcal{A} \to \mathbb{R}$ such that $h = g_0 + i g_1$, and in this case the integral of h with respect to μ is defined as

$$\int h(x) \, \mathrm{d}\mu(x) = \int g_0(x) \, \mathrm{d}\mu(x) + i \int g_1(x) \, \mathrm{d}\mu(x). \tag{1.213}$$

4. *Arbitrary Borel functions.* An arbitrary Borel function $f : \mathcal{A} \to \mathcal{B}$ is said to be *integrable* with respect to a given measure $\mu : \mathrm{Borel}(\mathcal{A}) \to [0, \infty]$ if and only if there exists a basis $\{w_1, \ldots, w_m\}$ of $\mathcal{W}$, for $\mathcal{W}$ being the finite-dimensional vector space for which $\mathcal{A} \subseteq \mathcal{W}$, along with integrable functions $g_1, \ldots, g_m : \mathcal{A} \to \mathbb{R}$ or $g_1, \ldots, g_m : \mathcal{A} \to \mathbb{C}$ (depending on whether $\mathcal{W}$ is a real or complex vector space) such that

$$f(x) = \sum_{k=1}^{m} g_k(x) w_k. \tag{1.214}$$

In this case, the integral of f with respect to μ is defined as

$$\int f(x)\,\mathrm{d}\mu(x) = \sum_{k=1}^{m} \left(\int g_k(x)\,\mathrm{d}\mu(x) \right) w_k. \tag{1.215}$$

The fact that the third and fourth items in this list lead to uniquely defined integrals of integrable functions is not immediate and requires a proof.

A selection of properties and conventions regarding integrals defined in this way, targeted to the specific needs of this book, follows.

1. *Linearity*. For integrable functions f and g, and scalar values α and β, one has
$$\int (\alpha f(x) + \beta g(x))\,\mathrm{d}\mu(x) = \alpha \int f(x)\,\mathrm{d}\mu(x) + \beta \int g(x)\,\mathrm{d}\mu(x). \tag{1.216}$$

2. *Standard Borel measure as the default*. Hereafter in this book, whenever $f : \mathbb{R} \to \mathbb{R}$ is an integrable function, and ν denotes the standard Borel measure on $\mathbb{R}$, the shorthand notation
$$\int f(\alpha)\,\mathrm{d}\alpha = \int f(\alpha)\,\mathrm{d}\nu(\alpha) \tag{1.217}$$
will be used. It is the case that, whenever f is an integrable function for which the commonly studied *Riemann integral* is defined, the Riemann integral will be in agreement with the integral defined as above for the standard Borel measure—so this shorthand notation is not likely to lead to confusion or ambiguity.

3. *Integration over subsets*. For an integrable function $f : \mathcal{A} \to \mathcal{B}$ and a Borel subset $\mathcal{C} \in \mathrm{Borel}(\mathcal{A})$, one defines
$$\int_{\mathcal{C}} f(x)\,\mathrm{d}\mu(x) = \int f(x)\chi_{\mathcal{C}}(x)\,\mathrm{d}\mu(x), \tag{1.218}$$
for $\chi_{\mathcal{C}}$ being the characteristic function of $\mathcal{C}$. The notation
$$\int_{\beta}^{\gamma} f(\alpha)\,\mathrm{d}\alpha = \int_{[\beta,\gamma]} f(\alpha)\,\mathrm{d}\alpha \tag{1.219}$$
is also used in the case that f takes the form $f : \mathbb{R} \to \mathcal{B}$ and $\beta, \gamma \in \mathbb{R}$ satisfy $\beta \leq \gamma$.

4. *Order of integration.* Suppose that $\mathcal{A}_0 \subseteq \mathcal{V}_0$, $\mathcal{A}_1 \subseteq \mathcal{V}_1$, and $\mathcal{B} \subseteq \mathcal{W}$ are subsets of finite-dimensional real or complex vector spaces, where it is to be assumed that $\mathcal{V}_0$ and $\mathcal{V}_1$ are either both real or both complex for simplicity. If $\mu_0 : \text{Borel}(\mathcal{A}_0) \to [0, \infty]$ and $\mu_1 : \text{Borel}(\mathcal{A}_0) \to [0, \infty]$ are Borel measures, $f : \mathcal{A}_0 \times \mathcal{A}_1 \to \mathcal{B}$ is a Borel function, and f is integrable with respect to the product measure $\mu_0 \times \mu_1$, then it holds (by a theorem known as *Fubini's theorem*) that

$$\begin{aligned} \int \biggl(\int f(x,y) \, \mathrm{d}\mu_0(x) \biggr) \mathrm{d}\mu_1(y) &= \int f(x,y) \, \mathrm{d}(\mu_0 \times \mu_1)(x,y) \\ &= \int \biggl(\int f(x,y) \, \mathrm{d}\mu_1(y) \biggr) \mathrm{d}\mu_0(x). \end{aligned} \tag{1.220}$$

Convex sets, cones, and functions

Let $\mathcal{V}$ be a vector space over the real or complex numbers. A subset $\mathcal{C}$ of $\mathcal{V}$ is *convex* if, for all vectors $u, v \in \mathcal{C}$ and scalars $\lambda \in [0, 1]$, it holds that

$$\lambda u + (1 - \lambda) v \in \mathcal{C}. \tag{1.221}$$

Intuitively speaking, this means that for any two distinct elements u and v of $\mathcal{C}$, the line segment whose endpoints are u and v lies entirely within $\mathcal{C}$. The intersection of any collection of convex sets is also convex.

If $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, either both over the real numbers or both over the complex numbers, and $\mathcal{A} \subseteq \mathcal{V}$ and $\mathcal{B} \subseteq \mathcal{W}$ are convex sets, then the set

$$\{u \oplus v \,:\, u \in \mathcal{A},\, v \in \mathcal{B}\} \subseteq \mathcal{V} \oplus \mathcal{W} \tag{1.222}$$

is also convex. Moreover, if $A \in \mathrm{L}(\mathcal{V}, \mathcal{W})$ is an operator, then the set

$$\{Au \,:\, u \in \mathcal{A}\} \subseteq \mathcal{W} \tag{1.223}$$

is convex as well.

A set $\mathcal{K} \subseteq \mathcal{V}$ is a *cone* if, for all choices of $u \in \mathcal{K}$ and $\lambda \geq 0$, one has that $\lambda u \in \mathcal{K}$. The cone *generated* by a set $\mathcal{A} \subseteq \mathcal{V}$ is defined as

$$\text{cone}(\mathcal{A}) = \bigl\{\lambda u \,:\, u \in \mathcal{A},\, \lambda \geq 0\bigr\}. \tag{1.224}$$

If $\mathcal{A}$ is a compact set that does not include 0, then $\text{cone}(\mathcal{A})$ is necessarily a closed set. A *convex cone* is simply a cone that is also convex. A cone $\mathcal{K}$ is

convex if and only if it is closed under addition, meaning that $u + v \in \mathcal{K}$ for every choice of $u, v \in \mathcal{K}$.

A function $f : \mathcal{C} \to \mathbb{R}$ defined on a convex set $\mathcal{C} \subseteq \mathcal{V}$ is a *convex function* if and only if the inequality

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v) \tag{1.225}$$

holds for all $u, v \in \mathcal{C}$ and $\lambda \in [0,1]$. A function $f : \mathcal{C} \to \mathbb{R}$ defined on a convex set $\mathcal{C} \subseteq \mathcal{V}$ is a *midpoint convex function* if and only if the inequality

$$f\Big(\frac{u+v}{2}\Big) \leq \frac{f(u) + f(v)}{2} \tag{1.226}$$

holds for all $u, v \in \mathcal{C}$. Every continuous midpoint convex function is necessarily convex.

A function $f : \mathcal{C} \to \mathbb{R}$ defined on a convex set $\mathcal{C} \subseteq \mathcal{V}$ is a *concave function* if and only if $-f$ is convex. Equivalently, f is concave if and only if the reverse of the inequality (1.225) holds for all $u, v \in \mathcal{C}$ and $\lambda \in [0,1]$. Similarly, a function $f : \mathcal{C} \to \mathbb{R}$ defined on a convex set $\mathcal{C} \subseteq \mathcal{V}$ is a *midpoint concave function* if and only if $-f$ is a midpoint convex function, and therefore every continuous midpoint concave function is concave.

Convex hulls

For any alphabet Σ, a vector $p \in \mathbb{R}^{\Sigma}$ is said to be a *probability vector* if it holds that $p(a) \geq 0$ for all $a \in \Sigma$ and

$$\sum_{a \in \Sigma} p(a) = 1. \tag{1.227}$$

The set of all such vectors will be denoted $\mathcal{P}(\Sigma)$.

For any vector space $\mathcal{V}$ and any subset $\mathcal{A} \subseteq \mathcal{V}$, a *convex combination* of vectors in $\mathcal{A}$ is any expression of the form

$$\sum_{a \in \Sigma} p(a) u_a, \tag{1.228}$$

for some choice of an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and a collection

$$\{u_a : a \in \Sigma\} \subseteq \mathcal{A} \tag{1.229}$$

of vectors in $\mathcal{A}$.

The *convex hull* of a set $\mathcal{A} \subseteq \mathcal{V}$, denoted $\operatorname{conv}(\mathcal{A})$, is the intersection of all convex sets containing $\mathcal{A}$. The set $\operatorname{conv}(\mathcal{A})$ is equal to the set of all vectors that may be written as a convex combination of elements of $\mathcal{A}$. (This is true even in the case that $\mathcal{A}$ is infinite.) The convex hull $\operatorname{conv}(\mathcal{A})$ of a closed set $\mathcal{A}$ need not itself be closed. However, if $\mathcal{A}$ is compact, then so too is $\operatorname{conv}(\mathcal{A})$.

The following theorem provides an upper bound on the number of elements over which one must take convex combinations in order to generate every point in the convex hull of a given set. The theorem refers to the notion of an *affine subspace*: a set $\mathcal{U} \subseteq \mathcal{V}$ is an affine subspace of $\mathcal{V}$ having dimension n if and only if there exists a subspace $\mathcal{W} \subseteq \mathcal{V}$ of dimension n and a vector $u \in \mathcal{V}$ such that

$$\mathcal{U} = \{u + v \,:\, v \in \mathcal{W}\}. \tag{1.230}$$

Theorem 1.9 (Carathéodory's theorem)**.** *Let $\mathcal{V}$ be a real vector space and let $\mathcal{A}$ be a subset of $\mathcal{V}$. Assume, moreover, that $\mathcal{A}$ is contained in an affine subspace of $\mathcal{V}$ having dimension n. For every vector $v \in \operatorname{conv}(\mathcal{A})$ in the convex hull of $\mathcal{A}$, there exist $m \leq n + 1$ vectors $u_1, \ldots, u_m \in \mathcal{A}$ such that $v \in \operatorname{conv}\{u_1, \ldots, u_m\}$.*

Extreme points

A point $w \in \mathcal{C}$ in a convex set $\mathcal{C}$ is said to be an *extreme point* of $\mathcal{C}$ if, for every expression

$$w = \lambda u + (1 - \lambda)v \tag{1.231}$$

for which $u, v \in \mathcal{C}$ and $\lambda \in (0, 1)$, it holds that $u = v = w$. In words, the extreme points are those elements of $\mathcal{C}$ that do not lie properly between two distinct points of $\mathcal{C}$.

The following theorem states that every convex and compact subset of a finite-dimensional vector space, over the real or complex numbers, is equal to the convex hull of its extreme points.

Theorem 1.10 (Minkowski)**.** *Let $\mathcal{V}$ be a finite-dimensional vector space over the real or complex numbers, let $\mathcal{C} \subseteq \mathcal{V}$ be a compact and convex set, and let $\mathcal{A} \subset \mathcal{C}$ be the set of extreme points of $\mathcal{C}$. It holds that $\mathcal{C} = \operatorname{conv}(\mathcal{A})$.*

A few examples of convex and compact sets, along with an identification of their extreme points, follow.

1. *The spectral norm unit ball.* For any complex Euclidean space $\mathcal{X}$, the set

$$\big\{X \in \mathrm{L}(\mathcal{X}) \,:\, \|X\| \leq 1\big\} \tag{1.232}$$

is a convex and compact set. The extreme points of this set are the unitary operators $\mathrm{U}(\mathcal{X})$.

2. *The trace-norm unit ball.* For any complex Euclidean space $\mathcal{X}$, the set

$$\big\{X \in \mathrm{L}(\mathcal{X}) \,:\, \|X\|_1 \leq 1\big\} \tag{1.233}$$

is a convex and compact set. The extreme points of this set are those operators of the form uv^* for $u, v \in \mathcal{X}$ unit vectors.

3. *Density operators.* For any complex Euclidean space $\mathcal{X}$, the set $\mathrm{D}(\mathcal{X})$ of density operators acting on $\mathcal{X}$ is convex and compact. The extreme points of $\mathrm{D}(\mathcal{X})$ coincide with the rank-one projection operators. These are the operators of the form uu^* for $u \in \mathcal{X}$ being a unit vector.

4. *Probability vectors.* For any alphabet Σ, the set of probability vectors $\mathcal{P}(\Sigma)$ is convex and compact. The extreme points of this set are the elements of the standard basis $\{e_a \,:\, a \in \Sigma\}$ of $\mathbb{R}^{\Sigma}$.

Hyperplane separation and min-max theorems

Convex sets in real Euclidean spaces possess a fundamentally important property: every vector lying outside of a given convex set in a real Euclidean space can be separated from that convex set by a *hyperplane*. That is, if the underlying real Euclidean space has dimension n, then there exists an affine subspace of that space having dimension $n - 1$ that divides the entire space into two half-spaces—one containing the convex set and the other containing the chosen point lying outside of the convex set. The following theorem represents one specific formulation of this fact.

Theorem 1.11 (Hyperplane separation theorem)**.** *Let $\mathcal{V}$ be a real Euclidean space, let $\mathcal{C} \subset \mathcal{V}$ be a closed, convex subset of $\mathcal{V}$, and let $u \in \mathcal{V}$ be a vector with $u \notin \mathcal{C}$. There exists a vector $v \in \mathcal{V}$ and a scalar $\alpha \in \mathbb{R}$ such that*

$$\langle v, u \rangle < \alpha \leq \langle v, w \rangle \tag{1.234}$$

for all $w \in \mathcal{C}$. If $\mathcal{C}$ is a cone, then v may be chosen so that (1.234) holds for $\alpha = 0$.

Another theorem concerning convex sets that finds uses in the theory of quantum information is the following theorem.

Theorem 1.12 (Sion's min-max theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be real or complex Euclidean spaces, let $\mathcal{A} \subseteq \mathcal{X}$ and $\mathcal{B} \subseteq \mathcal{Y}$ be convex sets with $\mathcal{B}$ compact, and let $f : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ be a continuous function such that*

1. *$x \mapsto f(x, y)$ is a concave function on $\mathcal{A}$ for all $y \in \mathcal{B}$, and*
2. *$y \mapsto f(x, y)$ is a convex function on $\mathcal{B}$ for all $x \in \mathcal{A}$.*

It holds that

$$\inf_{x \in \mathcal{A}} \max_{y \in \mathcal{B}} f(x, y) = \max_{y \in \mathcal{B}} \inf_{x \in \mathcal{A}} f(x, y). \tag{1.235}$$

1.2.2 Semidefinite programming

The paradigm of *semidefinite programming* finds numerous applications in the theory of quantum information, both analytical and computational. This section describes a formulation of semidefinite programming that is well-suited to its (primarily analytical) applications found in this book.

Definitions associated with semidefinite programs

Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a Hermiticity-preserving map, and let $A \in \mathrm{Herm}(\mathcal{X})$ and $B \in \mathrm{Herm}(\mathcal{Y})$ be Hermitian operators. A *semidefinite program* is a triple (Φ, A, B), with which the following pair of optimization problems is associated:

Primal problem		Dual problem	
maximize:	$\langle A, X \rangle$	minimize:	$\langle B, Y \rangle$
subject to:	$\Phi(X) = B$,	subject to:	$\Phi^*(Y) \geq A$,
	$X \in \mathrm{Pos}(\mathcal{X})$.		$Y \in \mathrm{Herm}(\mathcal{Y})$.

With these problems in mind, one defines the *primal feasible* set $\mathcal{A}$ and the *dual feasible* set $\mathcal{B}$ of (Φ, A, B) as follows:

$$\begin{aligned} \mathcal{A} &= \big\{X \in \mathrm{Pos}(\mathcal{X}) \,:\, \Phi(X) = B\big\}, \\ \mathcal{B} &= \big\{Y \in \mathrm{Herm}(\mathcal{Y}) \,:\, \Phi^*(Y) \geq A\big\}. \end{aligned} \tag{1.236}$$

Operators $X \in \mathcal{A}$ and $Y \in \mathcal{B}$ are also said to be *primal feasible* and *dual feasible*, respectively.

The function $X \mapsto \langle A, X \rangle$ from $\mathrm{Herm}(\mathcal{X})$ to $\mathbb{R}$ is called the *primal objective function*, and the function $Y \mapsto \langle B, Y \rangle$ from $\mathrm{Herm}(\mathcal{Y})$ to $\mathbb{R}$ is called the *dual objective function* of (Φ, A, B). The *optimum values* associated with the primal and dual problems are defined as

$$\alpha = \sup\bigl\{ \langle A, X \rangle \,:\, X \in \mathcal{A} \bigr\} \quad \text{and} \quad \beta = \inf\bigl\{ \langle B, Y \rangle \,:\, Y \in \mathcal{B} \bigr\}, \tag{1.237}$$

respectively. (If it is the case that $\mathcal{A} = \varnothing$ or $\mathcal{B} = \varnothing$, then one defines $\alpha = -\infty$ and $\beta = \infty$, respectively.)

Semidefinite programming duality

Semidefinite programs have associated with them a notion of *duality*, which refers to the special relationship between the primal and dual problems.

The property of *weak duality*, which holds for all semidefinite programs, is that the primal optimum can never exceed the dual optimum. In more succinct terms, it necessarily holds that $\alpha \leq \beta$. This implies that every dual feasible operator $Y \in \mathcal{B}$ provides an upper bound of $\langle B, Y \rangle$ on the value $\langle A, X \rangle$ that is achievable over all choices of a primal feasible $X \in \mathcal{A}$, and likewise every primal feasible operator $X \in \mathcal{A}$ provides a lower bound of $\langle A, X \rangle$ on the value $\langle B, Y \rangle$ that is achievable over all dual feasible operators $Y \in \mathcal{B}$.

It is not always the case that the primal optimum and dual optimum of a semidefinite program (Φ, A, B) agree, but for many semidefinite programs that arise naturally in applications, the primal optimum and dual optimum will be equal. This situation is called *strong duality*. The following theorem provides one set of conditions under which strong duality is guaranteed.

Theorem 1.13 (Slater's theorem for semidefinite programs)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a Hermiticity-preserving map, and let $A \in \mathrm{Herm}(\mathcal{X})$ and $B \in \mathrm{Herm}(\mathcal{Y})$ be Hermitian operators. Letting $\mathcal{A}$, $\mathcal{B}$, α, and β be as defined above for the semidefinite program (Φ, A, B), one has the following two implications:*

1. *If α is finite and there exists an operator $Y \in \mathrm{Herm}(\mathcal{Y})$ such that $\Phi^*(Y) > A$, then $\alpha = \beta$ and there exists $X \in \mathcal{A}$ such that $\langle A, X \rangle = \alpha$.*
2. *If β is finite and there exists an operator $X \in \mathrm{Pd}(\mathcal{X})$ such that $\Phi(X) = B$, then $\alpha = \beta$ and there exists $Y \in \mathcal{B}$ such that $\langle B, Y \rangle = \beta$.*

In the situation in which the optimum values of the primal and dual problems are equal, and are both achieved for some choice of feasible operators, a simple relationship between these operators must hold. The relationship is known as *complementary slackness*, and is expressed by the following proposition.

Proposition 1.14 (Complementary slackness for semidefinite programs)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a Hermiticity-preserving map, and let $A \in \mathrm{Herm}(\mathcal{X})$ and $B \in \mathrm{Herm}(\mathcal{Y})$ be Hermitian operators. Let $\mathcal{A}$ and $\mathcal{B}$ be the primal-feasible and dual-feasible sets associated with the semidefinite program (Φ, A, B), and suppose that $X \in \mathcal{A}$ and $Y \in \mathcal{B}$ satisfy $\langle A, X \rangle = \langle B, Y \rangle$. It holds that*

$$\Phi^*(Y)X = AX. \tag{1.238}$$

Simplified forms and alternative expressions of semidefinite programs

Semidefinite programs are typically presented in a way that is somewhat less formal than a precise specification of a triple (Φ, A, B), for $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ being a Hermiticity-preserving map and $A \in \mathrm{Herm}(\mathcal{X})$ and $B \in \mathrm{Herm}(\mathcal{Y})$ being Hermitian operators. Rather, the primal and dual problems are stated directly, often in a simplified form, and it is sometimes left to the reader to formulate a triple (Φ, A, B) that corresponds to the simplified problem statements.

Two examples of semidefinite programs follow, in both cases including their formal specifications and simplified forms.

Example 1.15 (Semidefinite program for the trace norm)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $K \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ be any operator. Define a Hermiticity-preserving map $\Phi \in \mathrm{T}(\mathcal{X} \oplus \mathcal{Y})$ as

$$\Phi \begin{pmatrix} X & \cdot \\ \cdot & Y \end{pmatrix} = \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \tag{1.239}$$

for all $X \in \mathrm{L}(\mathcal{X})$ and $Y \in \mathrm{L}(\mathcal{Y})$, where the dots represent elements of $\mathrm{L}(\mathcal{X}, \mathcal{Y})$ and $\mathrm{L}(\mathcal{Y}, \mathcal{X})$ that are effectively zeroed out by Φ. The map Φ is self-adjoint: $\Phi^* = \Phi$. Also define $A, B \in \mathrm{Herm}(\mathcal{X} \oplus \mathcal{Y})$ as

$$A = \frac{1}{2} \begin{pmatrix} 0 & K^* \\ K & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} \mathbb{1}_{\mathcal{X}} & 0 \\ 0 & \mathbb{1}_{\mathcal{Y}} \end{pmatrix}. \tag{1.240}$$

The primal and dual problems associated with the semidefinite program (Φ, A, B) may, after some simplifications, be expressed as follows:

Primal problem		Dual problem	
maximize:	$\frac{1}{2}\langle K, Z\rangle + \frac{1}{2}\langle K^*, Z^*\rangle$	minimize:	$\frac{1}{2}\operatorname{Tr}(X) + \frac{1}{2}\operatorname{Tr}(Y)$
subject to:	$\begin{pmatrix} \mathbb{1}_{\mathcal{X}} & Z^* \\ Z & \mathbb{1}_{\mathcal{Y}} \end{pmatrix} \geq 0,$	subject to:	$\begin{pmatrix} X & -K^* \\ -K & Y \end{pmatrix} \geq 0,$
	$Z \in \mathrm{L}(\mathcal{X}, \mathcal{Y}).$		$X \in \mathrm{Pos}(\mathcal{X}),$
			$Y \in \mathrm{Pos}(\mathcal{Y}).$

The primal and dual optima are equal for all choices of K, and given by $\|K\|_1$. (Given a singular value decomposition of K, one can construct both a primal feasible point and a dual feasible point achieving this value.)

A standard way of expressing this semidefinite program would be to list only the simplified primal and dual problems given above, letting the triple (Φ, A, B) be specified implicitly.

Example 1.16 (Semidefinite programs with inequality constraints)**.** Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Z})$ be Hermiticity-preserving maps, and let $A \in \mathrm{Herm}(\mathcal{X})$, $B \in \mathrm{Herm}(\mathcal{Y})$, and $C \in \mathrm{Herm}(\mathcal{Z})$ be Hermitian operators. Define a map

$$\Xi \in \mathrm{T}(\mathcal{X} \oplus \mathcal{Z}, \mathcal{Y} \oplus \mathcal{Z}) \tag{1.241}$$

as

$$\Xi \begin{pmatrix} X & \cdot \\ \cdot & Z \end{pmatrix} = \begin{pmatrix} \Phi(X) & 0 \\ 0 & \Psi(X) + Z \end{pmatrix} \tag{1.242}$$

for all $X \in \mathrm{L}(\mathcal{X})$ and $Z \in \mathrm{L}(\mathcal{Z})$. (Similar to the previous example, the dots in the argument to Ξ represent arbitrary elements of $\mathrm{L}(\mathcal{X}, \mathcal{Z})$ and $\mathrm{L}(\mathcal{Z}, \mathcal{X})$ upon which Ξ does not depend.) The adjoint map

$$\Xi^* \in \mathrm{T}(\mathcal{Y} \oplus \mathcal{Z}, \mathcal{X} \oplus \mathcal{Z}) \tag{1.243}$$

to Ξ is given by

$$\Xi^* \begin{pmatrix} Y & \cdot \\ \cdot & Z \end{pmatrix} = \begin{pmatrix} \Phi^*(Y) + \Psi^*(Z) & 0 \\ 0 & Z \end{pmatrix}. \tag{1.244}$$

The primal and dual problems of the semidefinite program specified by the map Ξ, together with the Hermitian operators

$$\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \in \mathrm{Herm}(\mathcal{X} \oplus \mathcal{Z}) \quad \text{and} \quad \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix} \in \mathrm{Herm}(\mathcal{Y} \oplus \mathcal{Z}), \tag{1.245}$$

may be simplified as follows:

Primal problem		Dual problem	
maximize:	$\langle A, X \rangle$	minimize:	$\langle B, Y \rangle + \langle C, Z \rangle$
subject to:	$\Phi(X) = B,$	subject to:	$\Phi^*(Y) + \Psi^*(Z) \geq A,$
	$\Psi(X) \leq C,$		$Y \in \mathrm{Herm}(\mathcal{Y}),$
	$X \in \mathrm{Pos}(\mathcal{X}).$		$Z \in \mathrm{Pos}(\mathcal{Z}).$

It is sometimes convenient to consider semidefinite programming problems of this form, that include both equality and inequality constraints in the primal problem, as opposed to just equality constraints.

1.2.3 Probability theory

Concepts from probability theory will play an important role throughout much of this book. Probability distributions over alphabets or other finite sets will be viewed as having fundamental importance; they arise naturally when information-theoretic tasks and settings are considered. The reader is assumed to have familiarity with basic probability theory for distributions over sets with finitely many elements. It will also be convenient to use the language of probability theory to discuss properties of Borel measures.

Random variables distributed with respect to probability measures

Suppose that $\mathcal{A}$ is a subset of a finite-dimensional real or complex vector space $\mathcal{V}$ and $\mu : \mathrm{Borel}(\mathcal{A}) \to [0, 1]$ is a probability measure (by which it is meant that μ is a normalized Borel measure). A *random variable* X distributed with respect to μ is a real-valued, integrable Borel function of the form

$$X : \mathcal{A} \to \mathbb{R}, \tag{1.246}$$

which is typically viewed as representing an outcome of a random process of some sort.

For every Borel subset $\mathcal{B} \subseteq \mathbb{R}$ of the real numbers, the probability that X takes a value in $\mathcal{B}$ is defined as

$$\Pr(X \in \mathcal{B}) = \mu\big(\big\{u \in \mathcal{A} \,:\, X(u) \in \mathcal{B}\big\}\big). \tag{1.247}$$

As a matter of notational convenience, one often writes expressions such as

$$\Pr(X \geq \beta) \quad \text{and} \quad \Pr(|X - \beta| \geq \varepsilon), \tag{1.248}$$

which are to be understood as meaning $\Pr(X \in \mathcal{B})$ for

$$\mathcal{B} = \{\alpha \in \mathbb{R} \,:\, \alpha \geq \beta\} \quad \text{and} \quad \mathcal{B} = \{\alpha \in \mathbb{R} \,:\, |\alpha - \beta| \geq \varepsilon\}, \tag{1.249}$$

respectively. Other expressions of this form are interpreted in an analogous way.

The *expected value* (or *mean value*) of a random variable X, distributed with respect to a probability measure $\mu : \text{Borel}(\mathcal{A}) \to [0, 1]$, is defined as

$$\mathrm{E}(X) = \int X(v)\, \mathrm{d}\mu(v). \tag{1.250}$$

If X is a random variable taking nonnegative real values, then it holds that

$$\mathrm{E}(X) = \int_0^\infty \Pr(X \geq \lambda)\, \mathrm{d}\lambda. \tag{1.251}$$

Random variables for discrete distributions

For a given alphabet Σ and a probability vector $p \in \mathcal{P}(\Sigma)$, one may also define a random variable X, distributed with respect to p, in an analogous way to a random variable distributed with respect to a Borel measure. In particular, such a random variable is a function of the form

$$X : \Sigma \to \mathbb{R}, \tag{1.252}$$

and for every subset $\Gamma \subseteq \Sigma$ one writes

$$\Pr(X \in \Gamma) = \sum_{a \in \Gamma} p(a). \tag{1.253}$$

In this case, the *expected value* (or *mean value*) of X is

$$\mathrm{E}(X) = \sum_{a \in \Sigma} p(a) X(a). \tag{1.254}$$

It is, in some sense, not necessary for random variables distributed with respect to probability vectors of the form $p \in \mathcal{P}(\Sigma)$ to be viewed as being fundamentally different from random variables distributed with respect to Borel probability measures. Indeed, one may consider the set

$$\{1, \ldots, n\} \subset \mathbb{R}, \tag{1.255}$$

for some choice of a positive integer n, and observe that every subset of $\{1, \ldots, n\}$ is a Borel subset of this set. The Borel probability measures

$$\mu : \text{Borel}(\{1, \ldots, n\}) \to [0, 1] \tag{1.256}$$

coincide precisely with the set of all probability vectors $p \in \mathcal{P}(\{1, \ldots, n\})$ through the equations

$$\mu(\mathcal{B}) = \sum_{b \in \mathcal{B}} p(b) \quad \text{and} \quad p(a) = \mu(\{a\}), \tag{1.257}$$

for every $\mathcal{B} \in \text{Borel}(\{1, \ldots, n\})$ and $a \in \{1, \ldots, n\}$.

Thus, by associating an arbitrary alphabet Σ with the set $\{1, \ldots, n\}$, one finds that a random variable distributed with respect to a probability vector $p \in \mathcal{P}(\Sigma)$ is represented by a random variable distributed with respect to a Borel probability measure.

Vector and operator valued random variables

It is sometimes convenient to define random variables that take vector or operator values, rather than real number values. Random variables of this sort will always be specified explicity in terms of ordinary random variables (i.e., ones that take real values) in this book. For example, given random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, for some choice of a positive integer n, one may refer to the vector-valued random variabes

$$(X_1, \ldots, X_n) \in \mathbb{R}^n \quad \text{and} \quad (X_1 + iY_1, \ldots, X_n + iY_n) \in \mathbb{C}^n. \tag{1.258}$$

The default meaning of the term *random variable* should be understood as referring to real-valued random variables, and the term *vector-valued random variable* will be used when referring to random variables obtained in the manner just described.

Independent and identically distributed random variables

Two random variables X and Y are said to be *independent* if and only if

$$\Pr((X, Y) \in \mathcal{A} \times \mathcal{B}) = \Pr(X \in \mathcal{A}) \Pr(Y \in \mathcal{B}) \tag{1.259}$$

for every choice of Borel subsets $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}$, and are said to be *identically distributed* if and only if

$$\Pr(X \in \mathcal{A}) = \Pr(Y \in \mathcal{A}) \tag{1.260}$$

for every Borel subset $\mathcal{A} \subseteq \mathbb{R}$. In general, these conditions do not require that X and Y are defined with respect to the same Borel measure. In both cases, these notions may be extended to more than two random variables, as well as to vector-valued random variables, in a straightforward way.

Suppose that $\mathcal{A}$ is a subset of a finite-dimensional real or complex vector space, $\mu : \text{Borel}(\mathcal{A}) \to [0, 1]$ is a probability measure, and $Y : \mathcal{A} \to \mathbb{R}$ is a random variable distributed with respect to μ. For any choice of a positive integer n, one may consider *independent and identically distributed* random variables $X_1, \ldots, X_n$, each being distributed in the same way as Y. For the purposes of this book, one may assume without a loss of generality that this means that $X_1, \ldots, X_n$ are Borel functions, taking the form

$$X_k : \mathcal{A}^n \to \mathbb{R} \tag{1.261}$$

and being defined as

$$X_k(u_1, \ldots, u_n) = Y(u_k) \tag{1.262}$$

for each k and each $(u_1, \ldots, u_n) \in \mathcal{A}^n$. Moreover, each X_k is understood to be distributed with respect to the n-fold product measure $\mu \times \cdots \times \mu$ on $\mathcal{A}^n$. In essence, this formal specification represents the simple and intuitive notion that $X_1, \ldots, X_n$ are uncorrelated copies of the random variable Y.

A few fundamental theorems

A few fundamental theorems concerning random variables will be used later in this book. While these theorem do hold for more general notions of random variables, the theorem statements that follow should be understood to apply to random variables distributed with respect to Borel probability measures (including random variables distributed with respect to probability vectors of the form $p \in \mathcal{P}(\Sigma)$ as a special case, as described above).

The first theorem to be stated in this subsection is *Markov's inequality*, which provides a sometimes coarse upper bound on the probability that a nonnegative random variable exceeds a given threshold value.

Theorem 1.17 (Markov's inequality)**.** *Let X be a random variable taking nonnegative real values, and let $\varepsilon > 0$ be a positive real number. It holds that*

$$\Pr\big(X \geq \varepsilon\big) \leq \frac{\mathrm{E}(X)}{\varepsilon}. \tag{1.263}$$

The next theorem, known as *Jensen's inequality*, concerns the expected value of a convex function applied to a random variable.

Theorem 1.18 (Jensen's inequality)**.** *Suppose that X is a random variable and $f : \mathbb{R} \to \mathbb{R}$ is a convex function. It holds that*

$$f\big(\mathrm{E}(X)\big) \leq \mathrm{E}(f(X)). \tag{1.264}$$

Two additional theorems—known as the *weak law of large numbers* and *Hoeffding's inequality*—provide bounds on the deviation of the average value of a collection of independent and identically distributed random variables from their mean value.

Theorem 1.19 (Weak law of large numbers)**.** *For every positive integer n, let X_n be a random variable having mean value α, and assume moreover that $X_1, \ldots, X_n$ are independent and identically distributed. For every positive real number $\varepsilon > 0$, it holds that*

$$\lim_{n\to\infty} \Pr\left(\left|\frac{X_1 + \cdots + X_n}{n} - \alpha\right| \geq \varepsilon\right) = 0. \tag{1.265}$$

Theorem 1.20 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_n$ be independent and identically distributed random variables taking values in the interval $[0, 1]$ and having mean value α. For every positive real number $\varepsilon > 0$ it holds that*

$$\Pr\left(\left|\frac{X_1 + \cdots + X_n}{n} - \alpha\right| \geq \varepsilon\right) \leq 2\exp\big(-2n\varepsilon^2\big). \tag{1.266}$$

Gaussian measure and normally distributed random variables

The *standard Gaussian measure* on $\mathbb{R}$ is the Borel probability measure

$$\gamma : \text{Borel}(\mathbb{R}) \to [0, 1] \tag{1.267}$$

defined as

$$\gamma(\mathcal{A}) = \frac{1}{\sqrt{2\pi}} \int_{\mathcal{A}} \exp\left(-\frac{\alpha^2}{2}\right) \mathrm{d}\alpha \tag{1.268}$$

for every $\mathcal{A} \in \text{Borel}(\mathbb{R})$, where the integral is to be taken with respect to the standard Borel measure on $\mathbb{R}$. The fact that this is a well-defined measure follows from the observation that the function

$$\alpha \mapsto \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2}\right) & \text{if } \alpha \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases} \tag{1.269}$$

is an integrable Borel function for every Borel subset $\mathcal{A} \subset \mathbb{R}$, and the fact that it is a probability measure follows from the Gaussian integral

$$\int \exp\left(-\frac{\alpha^2}{2}\right) \mathrm{d}\alpha = \sqrt{2\pi}. \tag{1.270}$$

A random variable X is a *standard normal random variable* if it is given by the identity function $X(\alpha) = \alpha$ and distributed with respect to the standard Gaussian measure γ on $\mathbb{R}$. This is equivalent to X being given by the function

$$X(\alpha) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2}\right) \tag{1.271}$$

and being distributed with respect to the standard Borel measure on $\mathbb{R}$—although the description in terms of the standard Gaussian measure will be the preferred description in this book.

The following integrals are among many integrals of a similar sort that are useful when reasoning about standard normal random variables.

1. For every positive real number $\lambda > 0$ and every real number $\beta \in \mathbb{R}$ it holds that
$$\int \exp\left(-\lambda\alpha^2 + \beta\alpha\right) \mathrm{d}\alpha = \sqrt{\frac{\pi}{\lambda}} \exp\left(\frac{\beta^2}{4\lambda}\right). \tag{1.272}$$

2. For every positive integer n, it holds that it holds that

$$\int_0^\infty \alpha^n \, \mathrm{d}\gamma(\alpha) = \frac{2^{\frac{n}{2}} \Gamma\left(\frac{n+1}{2}\right)}{2\sqrt{\pi}}, \tag{1.273}$$

where the Γ-function may be defined at positive half-integer points as follows:

$$\Gamma\left(\frac{n+1}{2}\right) = \begin{cases} \sqrt{\pi} & \text{if } n = 0 \\ 1 & \text{if } n = 1 \\ \frac{n-1}{2} \, \Gamma\left(\frac{n-1}{2}\right) & \text{if } n \geq 2. \end{cases} \tag{1.274}$$

3. For every positive real number $\lambda > 0$ and every pair of real numbers $\beta_0, \beta_1 \in \mathbb{R}$ with $\beta_0 \leq \beta_1$ it holds that

$$\int_{\beta_0}^{\beta_1} \alpha \exp(-\lambda \alpha^2) \, \mathrm{d}\alpha = \frac{1}{2\lambda} \exp\left(-\lambda \beta_0^2\right) - \frac{1}{2\lambda} \exp\left(-\lambda \beta_1^2\right). \tag{1.275}$$

This formula also holds for infinite values of β_0 and β_1, with the natural interpretation $\exp(-\infty) = 0$.

For every positive integer n, the *standard Gaussian measure* on $\mathbb{R}^n$ is the Borel probability measure

$$\gamma_n : \mathrm{Borel}(\mathbb{R}^n) \to [0, 1] \tag{1.276}$$

obtained by taking the n-fold product measure of γ with itself. Equivalently,

$$\gamma_n(\mathcal{A}) = (2\pi)^{-\frac{n}{2}} \int_{\mathcal{A}} \exp\left(-\frac{\|v\|^2}{2}\right) \mathrm{d}\nu_n(v), \tag{1.277}$$

where ν_n denotes the n-fold product measure of the standard Borel measure ν with itself.

The standard Gaussian measure is invariant under orthogonal transformations (which include rotations):

$$\gamma_n(U\mathcal{A}) = \gamma_n(\mathcal{A}) \tag{1.278}$$

for every Borel set $\mathcal{A} \subseteq \mathbb{R}^n$ and every orthogonal operator $U \in \mathrm{L}(\mathbb{R}^n)$, meaning one that satisfies $UU^{\mathrm{T}} = \mathbb{1}$. Thus, for independent and identically

distributed standard normal random variables $X_1, \ldots, X_n$, one has that the vector valued random variable $(X_1, \ldots, X_n)$ is identically distributed to the vector-valued random variable $(Y_1, \ldots, Y_n)$ obtained by defining

$$Y_k = \sum_{j=1}^{n} U(k, j) X_j \tag{1.279}$$

for each $k \in \{1, \ldots, n\}$, for $U \in \mathrm{L}(\mathbb{R}^n)$ being any orthogonal operator. As a consequence of this fact, one has that if the standard Gaussian measure is projected onto a subspace, it is equivalent to the standard Gaussian measure on that subspace.

Proposition 1.21. *Let m and n be positive integers satisfying $m < n$ and let $V \in \mathrm{L}(\mathbb{R}^m, \mathbb{R}^n)$ satisfy $V^{\mathsf{T}} V = \mathbb{1}$. For every Borel set $\mathcal{A} \subseteq \mathbb{R}^m$, one has*

$$\gamma_m(\mathcal{A}) = \gamma_n\big(\{u \in \mathbb{R}^n \,:\, V^{\mathsf{T}} u \in \mathcal{A}\}\big). \tag{1.280}$$

It follows from this proposition that the standard Gaussian measure $\gamma_n(\mathcal{V})$ of any proper subspace $\mathcal{V}$ of $\mathbb{R}^n$ is zero.

Finally, let $X_1, \ldots, X_n$ be independent and identically distributed standard normal random variables, and define a random variable

$$Y = \sqrt{X_1^2 + \cdots + X_n^2}. \tag{1.281}$$

The distribtion of Y is known as the *χ-distribution*. The mean value of Y has the following closed-form expression:

$$\mathrm{E}(Y) = \frac{\sqrt{2}\,\Gamma\big(\frac{n+1}{2}\big)}{\Gamma\big(\frac{n}{2}\big)}. \tag{1.282}$$

From this expression, it may be proved that

$$\mathrm{E}(Y) = v_n \sqrt{n}, \tag{1.283}$$

where $(v_1, v_2, \ldots)$ is a strictly increasing sequence that begins

$$v_1 = \sqrt{\frac{2}{\pi}}, \quad v_2 = \frac{\sqrt{\pi}}{2}, \quad v_3 = \sqrt{\frac{8}{3\pi}}, \quad \ldots \tag{1.284}$$

and converges to 1 in the limit as n goes to infinity.

1.3 Suggested references

Several textbooks cover the material on linear algebra summarized in this chapter; the classic books of Halmos [89] and Hoffman and Kunze [103] are two examples. Readers interested in a more modern development of linear algebra for finite dimensional spaces are referred to the book of Axler [19]. The books of Horn and Johnson [114] and Bhatia [41] also cover much of the material on linear algebra that has been summarized in this chapter (and a great deal more, including relevant theorems to be proved in subsequent chapters of this book), with a focus on the matrix-theoretic aspects of the subject.

There are also many textbooks on mathematical analysis, including the classic texts of Rudin [178] and Apostol [11], as well as the books of Bartle [24] and Halmos [88] that focus on measure theory. The book of Rockafellar [175] is a standard reference on convex analysis, and semidefinite programming is discussed in the book of Wolkowicz, Saigal, and Vandenberge [231]. The two volume collection of Feller [72, 73] is a standard reference on probability theory.

Chapter 2

Basic notions of quantum information

This chapter introduces the elementary notions of quantum information theory, including *registers*, *states*, *channels*, and *measurements*, that form the foundation upon which the theory of quantum information is built.

2.1 Registers and states

This first section of the chapter concerns *registers* and *states*. A register is an abstraction of a physical device in which quantum information may be stored, and the state of a register represents a description of its contents at a particular instant.

2.1.1 Registers and classical state sets

The term *register* is intended to be suggestive of a computer component in which some finite amount of data can be stored and manipulated. While this is a reasonable picture to keep in mind, it should be understood that any physical system in which a finite amount of data may be stored, and whose state may change over time, could be modeled as a register. For example, a register could represent a medium used to transmit information from a sender to a receiver. At an intuitive level, what is most important is that registers represent mathematical abstractions of physical objects, or parts of a physical objects, that store information.

Definition of registers

The following formal definition of a register is intended to capture a basic but nevertheless important idea, which is that multiple registers may be viewed collectively as forming a single register. It is natural to choose an inductive definition for this reason.

Definition 2.1. A *register* X is either one of the following two objects:

1. An alphabet Σ.
2. An n-tuple $\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$, where n is a positive integer and $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ are registers.

Registers of the first type are called *simple registers* and registers of the second type are called *compound registers* when it is helpful to distinguish them.

In the case of a simple register $\mathsf{X} = \Sigma$, the alphabet Σ represents the set of *classical states* that the register may store. The classical state set associated with a compound register will be specified shortly. As is suggested by the definition, registers will be denoted by capital letters in a *sans serif* font, such as X, Y, and Z. Sometimes registers will be subscripted, such as $\mathsf{X}_1, \ldots, \mathsf{X}_n$, when it is necessary to refer to a variable number of registers or convenient to name them in this way for some other reason.

Based on Definition 2.1, one may naturally identify a tree structure with a given register, with each leaf node corresponding to a simple register. A register Y is said to be a *subregister* of X if the tree associated with Y is a subtree of the tree associated with X.

Example 2.2. One may define registers X, Y_0, Y_1, Z_1, Z_2, and Z_3, as follows:

$$\begin{array}{lll} \mathsf{X} = (\mathsf{Y}_0, \mathsf{Y}_1), & \mathsf{Y}_0 = \{1,2,3,4\}, & \mathsf{Z}_1 = \{0,1\}, \\ & \mathsf{Y}_1 = (\mathsf{Z}_1, \mathsf{Z}_2, \mathsf{Z}_3), & \mathsf{Z}_2 = \{0,1\}, \\ & & \mathsf{Z}_3 = \{0,1\}. \end{array} \tag{2.1}$$

The tree associated with the register X is illustrated in Figure 2.1. The subregisters of X include Y_0, Y_1, Z_1, Z_2, Z_3, and (trivially) X itself.

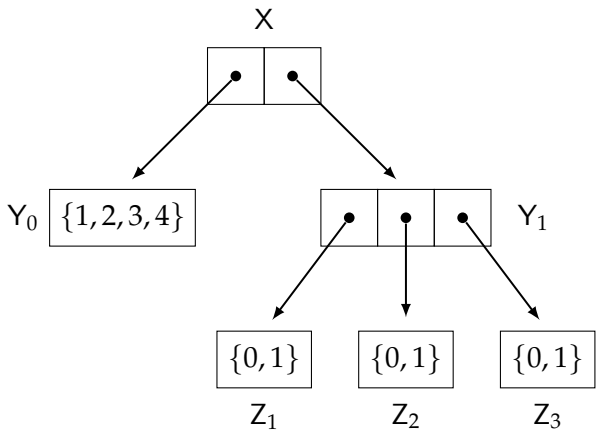


Figure 2.1: The tree associated with the registers described in Example 2.2.

The classical state set of a register

Every register has associated with it a *classical state set*, as specified by the following definition.

Definition 2.3. The *classical state set* of a register X is an alphabet, determined as follows.

1. If $\mathsf{X} = \Sigma$ is a simple register, the classical state set of X is Σ.
2. If $\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ is a compound register, the classical state set of X is the Cartesian product
$$\Sigma = \Gamma_1 \times \cdots \times \Gamma_n, \tag{2.2}$$
where Γ_k denotes the classical state set associated with the register Y_k for each $k \in \{1, \ldots, n\}$.

Elements of a register's classical state set are called *classical states* of that register.

The term *classical state* is intended to be suggestive of the classical notion of a state in computer science. Intuitively speaking, a classical state of a register can be recognized unambiguously, like the values 0 and 1 stored by a single bit memory component. The term *classical state* should not be confused with the term *state*, which by default will mean *quantum state* rather than *classical state* throughout this book.

A register is said to be *trivial* if its classical state set contains just a single element. Trivial registers are useless from an information-processing viewpoint, but mathematically it is convenient to allow for this possibility. The reader will note, however, that registers with *empty* classical state sets are disallowed by the definition. This is consistent with the intuition that registers represent physical systems; while it is possible that a physical system could have just one possible classical state, it is nonsensical for a system to have no states whatsoever.

Reductions of classical states

There is a straightforward way in which each classical state of a register uniquely determines a classical state for each of its subregisters. To be more precise, suppose that

$$\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n) \tag{2.3}$$

is a compound register. Let $\Gamma_1, \ldots, \Gamma_n$ denote the classical state sets of the registers $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$, respectively, so that the classical state set of X is equal to $\Sigma = \Gamma_1 \times \cdots \times \Gamma_n$. A given classical state $a = (b_1, \ldots, b_n)$ of X then determines that the classical state of Y_k is $b_k \in \Gamma_k$, for each $k \in \{1, \ldots, n\}$. By applying this definition recursively, one defines a unique classical state of each subregister of X.

Conversely, the classical state of any register is uniquely determined by the classical states of its simple subregisters. Every classical state of a given register X therefore uniquely determines a classical state of any register whose simple subregisters form a subset of those of X. For instance, if X takes the form (2.3), then one may wish to consider a new register

$$\mathsf{Z} = (\mathsf{Y}_{k_1}, \ldots, \mathsf{Y}_{k_m}) \tag{2.4}$$

for some choice of indices $1 \leq k_1 < \cdots < k_m \leq n$ (for instance). If $a = (b_1, \ldots, b_n)$ is the classical state of X at a particular moment, then the corresponding state of Z is $(b_{k_1}, \ldots, b_{k_m})$.

2.1.2 Quantum states of registers

Quantum states, as they will be presented in this book, may be viewed as being analogous to probabilistic states, with which the reader is assumed to have some familiarity.

A *probabilistic state* of a register X refers to a random mixture, or probability distribution, over the classical states of that register. Assuming the classical state set of X is Σ, a probabilistic state of X is identified with a probability vector $p \in \mathcal{P}(\Sigma)$, with the value $p(a)$ representing the probability associated with each classical state $a \in \Sigma$. It is typical that one views a probabilistic state as being a mathematical representation of a register's contents, or of a hypothetical individual's knowledge of that register's contents, at a particular moment.

The difference between probabilistic states and quantum states is that, whereas probabilistic states are represented by probability vectors, quantum states are represented by *density operators* (q.v. Section 1.1.2). Unlike the notion of a probabilistic state, which has a relatively clear and intuitive meaning, the notion of a quantum state can seem non-intuitive. While it is both natural and interesting to seek an understanding of why Nature appears to be well-modeled by quantum states in certain regimes, this book will not attempt to provide such an understanding: quantum states will be considered as mathematical objects and nothing more.

The complex Euclidean space associated with a register

It is helpful to introduce the following terminology to discuss quantum states in mathematical terms: the complex Euclidean space associated with a register X is defined to be $\mathbb{C}^{\Sigma}$, where Σ is the classical state set of X.

The complex Euclidean space associated with a given register will generally be denoted by the same letter as the register itself, but with a *scripted* font rather than a *sans serif* font. For example, the complex Euclidean space associated with a register X will be denoted $\mathcal{X}$, and the spaces associated with registers $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ will be denoted $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$. This association should be considered as a default assumption, and will not usually be mentioned explicitly when it is made.

The reader will note that the complex Euclidean space $\mathcal{X}$ associated with a compound register $\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ is given by the tensor product

$$\mathcal{X} = \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n. \tag{2.5}$$

This fact follows directly from the definition stating that the classical state set of X is given by $\Sigma = \Gamma_1 \times \cdots \times \Gamma_n$, assuming that the classical state sets of $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ are $\Gamma_1, \ldots, \Gamma_n$, respectively; one has that the complex Euclidean

space associated with X is

$$\mathcal{X} = \mathbb{C}^{\Sigma} = \mathbb{C}^{\Gamma_1 \times \cdots \times \Gamma_n} = \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n \tag{2.6}$$

for $\mathcal{Y}_1 = \mathbb{C}^{\Gamma_1}, \ldots, \mathcal{Y}_n = \mathbb{C}^{\Gamma_n}$.

Definition of quantum states

As stated above, quantum states are represented by density operators. The following definition makes this precise.

Definition 2.4. A *quantum state* is a density operator of the form $\rho \in \mathrm{D}(\mathcal{X})$ for some choice of a complex Euclidean space $\mathcal{X}$.

When one refers to a quantum state of a register X, it is to be understood that the state in question takes the form $\rho \in \mathrm{D}(\mathcal{X})$ for $\mathcal{X}$ being the complex Euclidean space associated with X. It is common that the term *state* is used in place of *quantum state* in the setting of quantum information, because it is the default assumption that one is primarily concerned with quantum states (as opposed to classical states and probabilistic states) in this setting.

Convex combinations of quantum states

For every complex Euclidean space $\mathcal{X} = \mathbb{C}^{\Sigma}$, the set $\mathrm{D}(\mathcal{X})$ is a convex set. For any choice of an alphabet Γ, a collection

$$\{\rho_a \,:\, a \in \Gamma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{2.7}$$

of quantum states, and a probability vector $p \in \mathcal{P}(\Gamma)$, it therefore holds that the convex combination

$$\rho = \sum_{a \in \Gamma} p(a)\rho_a \tag{2.8}$$

is an element of $\mathrm{D}(\mathcal{X})$. The state ρ defined by the equation (2.8) is said to be a *mixture* of the states $\{\rho_a \,:\, a \in \Gamma\}$ according to the probability vector p. Supposing that X is a register whose associated complex Euclidean space is $\mathcal{X}$, it will be taken as an axiom that a random selection of $a \in \Gamma$ according to the probability vector p, followed by a preparation of X in the state ρ_a, results in X being in the state ρ defined in (2.8). More succinctly, random selections of quantum states are represented as convex combinations of density operators.

The notion of a probability distribution over a finite set of quantum states arises frequently in the theory of quantum information. A distribution of the form described above may be succinctly represented by a function $\eta : \Gamma \to \mathrm{Pos}(\mathcal{X})$ satisfying the constraint

$$\mathrm{Tr}\left(\sum_{a\in\Gamma} \eta(a)\right) = 1. \tag{2.9}$$

A function η of this sort is called an *ensemble* of states. The interpretation of an ensemble of states $\eta : \Gamma \to \mathrm{Pos}(\mathcal{X})$ is that, for each element $a \in \Gamma$, the operator $\eta(a)$ represents a state together with the probability associated with that state: the probability is $\mathrm{Tr}(\eta(a))$, while the state is

$$\rho_a = \frac{\eta(a)}{\mathrm{Tr}(\eta(a))}. \tag{2.10}$$

(The operator ρ_a is, of course, determined only when $\eta(a) \neq 0$. In the case that $\eta(a) = 0$ for some choice of a, one does not generally need to specify a specific density operator ρ_a, as it corresponds to a discrete event that occurs with probability zero.)

Pure states

A state $\rho \in \mathrm{D}(\mathcal{X})$ is said to be a *pure state* if and only if it has rank equal to 1. Equivalently, ρ is a pure state if and only if there exists a unit vector $u \in \mathcal{X}$ such that

$$\rho = uu^*. \tag{2.11}$$

It follows from the spectral theorem (Corollary 1.4) that every quantum state is a mixture of pure quantum states, and moreover that the extreme points of the set $\mathrm{D}(\mathcal{X})$ are precisely the pure states of X.

It is common that one refers to the pure state (2.11) simply as u, rather than uu^*. There is an ambiguity that arises in following this convention: if one considers two unit vectors u and $v = \alpha u$, for any choice of $\alpha \in \mathbb{C}$ with $|\alpha| = 1$, then their corresponding pure states uu^* and vv^* are equal, as

$$vv^* = |\alpha|^2 uu^* = uu^*. \tag{2.12}$$

Fortunately, this convention does not generally cause confusion—it must simply be kept in mind that every pure state corresponds to an equivalence

class of unit vectors, where u and v are equivalent if and only if $v = \alpha u$ for some choice of $\alpha \in \mathbb{C}$ with $|\alpha| = 1$, and that any particular unit vector may be viewed as being a representative of a pure state from this equivalence class.

Flat states

A density operator $\rho \in \mathrm{D}(\mathcal{X})$ is said to be a *flat state* if and only if it holds that

$$\rho = \frac{\Pi}{\mathrm{Tr}(\Pi)} \tag{2.13}$$

for $\Pi \in \mathrm{Proj}(\mathcal{X})$ being a nonzero projection operator. The symbol ω will often be used to denote a flat state, and the notation

$$\omega_{\mathcal{V}} = \frac{\Pi_{\mathcal{V}}}{\mathrm{Tr}(\Pi_{\mathcal{V}})} \tag{2.14}$$

is sometimes used to denote the flat state proportional to the projection $\Pi_{\mathcal{V}}$ onto a nonzero subspace $\mathcal{V} \subseteq \mathcal{X}$. Specific examples of flat states include pure states, which correspond to the case that Π is a rank-one projection, and the *completely mixed state*

$$\omega = \frac{\mathbb{1}_{\mathcal{X}}}{\dim(\mathcal{X})}. \tag{2.15}$$

Intuitively speaking, the completely mixed state represents a state of ignorance, analogous to a uniform probability distribution over a given classical state set.

Classical states and probabilistic states as quantum states

Suppose that X is a register having classical state set Σ, so that the complex Euclidean space associated with X is $\mathcal{X} = \mathbb{C}^{\Sigma}$. Within the set $\mathrm{D}(\mathcal{X})$ of states of X, one may choose to represent the possible classical states Σ of X in the following simple way: the operator $E_{a,a} \in \mathrm{D}(\mathcal{X})$ is taken as a representation of the register X being in the classical state a, for each $a \in \Sigma$. Through this association, probabilistic states of registers correspond to diagonal density operators, with each probabilistic state $p \in \mathcal{P}(\Sigma)$ being represented by the density operator

$$\sum_{a \in \Sigma} p(a) E_{a,a} = \mathrm{Diag}(p). \tag{2.16}$$

In this way, the set of probabilistic states of a given register form a subset of the set of all quantum states of that register (with the containment being proper unless the register is trivial).[1]

Within certain contexts, it may be necessary or appropriate to specify that one or more registers are *classical registers*. Informally speaking, a classical register is one whose state is always a diagonal density operator, corresponding to a classical (probabilistic) state. A more formal and precise meaning of this terminology must be postponed until the section on quantum channels following this one.

Product states

Suppose $\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ is a compound register. A state $\rho \in \mathrm{D}(\mathcal{X})$ is said to be a *product state* of X if and only it takes the form

$$\rho = \sigma_1 \otimes \cdots \otimes \sigma_n \tag{2.17}$$

for $\sigma_1 \in \mathrm{D}(\mathcal{Y}_1), \ldots, \sigma_n \in \mathrm{D}(\mathcal{Y}_n)$ being states of $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$, respectively. Product states represent independence among the states of registers. It is sometimes said that the registers $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ are *independent* when the compound register $\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ is in a product state ρ of the form (2.17). When it is not the case that $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ are independent, they are said to be *correlated*.

Example 2.5. Consider a compound register of the form $\mathsf{X} = (\mathsf{Y}, \mathsf{Z})$, for Y and Z being registers sharing the classical state set $\{0, 1\}$. (Registers having the classical state set $\{0, 1\}$ are typically called *qubits*, which is short for *quantum bits*.)

The state $\rho \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$ defined as

$$\rho = \frac{1}{4} E_{0,0} \otimes E_{0,0} + \frac{1}{4} E_{0,0} \otimes E_{1,1} + \frac{1}{4} E_{1,1} \otimes E_{0,0} + \frac{1}{4} E_{1,1} \otimes E_{1,1} \tag{2.18}$$

is an example of a product state, as one may write

$$\rho = \left(\frac{1}{2} E_{0,0} + \frac{1}{2} E_{1,1} \right) \otimes \left(\frac{1}{2} E_{0,0} + \frac{1}{2} E_{1,1} \right). \tag{2.19}$$

[1] The other basic notions of quantum information to be discussed in this chapter have a similar character of admitting analogous probabilistic notions as special cases. In general, the theory of quantum information may be seen as an extension of classical information theory, including the study of random processes, protocols, and computations.

Equivalently, in matrix form, one has

$$\rho = \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \otimes \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \tag{2.20}$$

The states $\sigma, \tau \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$ defined as

$$\sigma = \frac{1}{2} E_{0,0} \otimes E_{0,0} + \frac{1}{2} E_{1,1} \otimes E_{1,1} \tag{2.21}$$

and

$$\tau = \frac{1}{2} E_{0,0} \otimes E_{0,0} + \frac{1}{2} E_{0,1} \otimes E_{0,1} + \frac{1}{2} E_{1,0} \otimes E_{1,0} + \frac{1}{2} E_{1,1} \otimes E_{1,1} \tag{2.22}$$

are examples of states that are not product states, as they cannot be written as tensor products, and therefore represent correlations between the registers Y and Z. In matrix form, these states are as follows:

$$\sigma = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}. \tag{2.23}$$

The states ρ and σ are diagonal, and therefore correspond to probabilistic states; ρ represents the situation in which Y and Z store independent random bits, while σ represents the situation in which Y and Z store perfectly correlated random bits. The state τ does not represent a probabilistic state, and more specifically is an example of an *entangled* state. Entanglement is a particular type of correlation having great significance in quantum information theory, and is the primary focus of Chapter 6.

Bases of density operators

It is an elementary, but nevertheless useful, fact that for every complex Euclidean space $\mathcal{X}$ there exist spanning sets of the space $\mathrm{L}(\mathcal{X})$ consisting only of density operators. One implication of this fact is that every linear mapping of the form

$$\phi : \mathrm{L}(\mathcal{X}) \to \mathbb{C} \tag{2.24}$$

is uniquely determined by its action on the elements of $\mathrm{D}(\mathcal{X})$. (This will imply, for instance, that channels and measurements are uniquely determined by their actions on density operators.) The following example describes one way of constructing such a spanning set.

Example 2.6. Let Σ be an alphabet, and assume that a total ordering has been defined on Σ. For every pair $(a, b) \in \Sigma \times \Sigma$, define a density operator $\rho_{a,b} \in \mathrm{D}\big(\mathbb{C}^{\Sigma}\big)$ as follows:

$$\rho_{a,b} = \begin{cases} E_{a,a} & \text{if } a = b \\ \frac{1}{2}(e_a + e_b)(e_a + e_b)^* & \text{if } a < b \\ \frac{1}{2}(e_a + ie_b)(e_a + ie_b)^* & \text{if } a > b. \end{cases} \tag{2.25}$$

For each pair $(a, b) \in \Sigma \times \Sigma$ with $a < b$, one has

$$\begin{aligned} \left(\rho_{a,b} - \frac{1}{2}\rho_{a,a} - \frac{1}{2}\rho_{b,b}\right) - i\left(\rho_{b,a} - \frac{1}{2}\rho_{a,a} - \frac{1}{2}\rho_{b,b}\right) &= E_{a,b}, \\ \left(\rho_{a,b} - \frac{1}{2}\rho_{a,a} - \frac{1}{2}\rho_{b,b}\right) + i\left(\rho_{b,a} - \frac{1}{2}\rho_{a,a} - \frac{1}{2}\rho_{b,b}\right) &= E_{b,a}, \end{aligned} \tag{2.26}$$

and from these equations it follows that $\operatorname{span}\{\rho_{a,b} : (a, b) \in \Sigma \times \Sigma\} = \mathrm{L}(\mathcal{X})$.

2.1.3 Reductions and purifications of quantum states

One may consider a register that is formed by removing one or more subregisters from a given compound register. The quantum state of any register that results from this process, viewed in isolation from the subregisters that were removed, is always uniquely determined by the state of the original compound register. This section describes how these states are determined, and further develops an important special case in which the state of the original compound register is pure.

The partial trace and reductions of quantum states

Let $\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ be a compound register, for $n \geq 2$. For any choice of $k \in \{1, \ldots, n\}$, one may form a new register

$$(\mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n) \tag{2.27}$$

by removing the register Y_k from X and leaving the remaining registers untouched. For every state $\rho \in \mathrm{D}(\mathcal{X})$ of X, the state of the register (2.27) that is determined by this process is called the *reduction* of ρ to the register (2.27), and is denoted $\rho[\mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n]$.

Specifically, this state is defined as

$$\rho[\mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n] = \mathrm{Tr}_{\mathcal{Y}_k}(\rho), \tag{2.28}$$

where

$$\mathrm{Tr}_{\mathcal{Y}_k} \in \mathrm{T}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_{k-1} \otimes \mathcal{Y}_{k+1} \otimes \cdots \otimes \mathcal{Y}_n) \tag{2.29}$$

denotes the *partial trace* mapping (q.v. Section 1.1.2).[2] This is the unique linear mapping that satisfies the equation

$$\mathrm{Tr}_{\mathcal{Y}_k}(Y_1 \otimes \cdots \otimes Y_n) = \mathrm{Tr}(Y_k)\, Y_1 \otimes \cdots \otimes Y_{k-1} \otimes Y_{k+1} \otimes \cdots \otimes Y_n \tag{2.30}$$

for all operators $Y_1 \in \mathrm{L}(\mathcal{Y}_1), \ldots, Y_n \in \mathrm{L}(\mathcal{Y}_n)$. Alternately, one may define

$$\mathrm{Tr}_{\mathcal{Y}_k} = \mathbb{1}_{\mathrm{L}(\mathcal{Y}_1)} \otimes \cdots \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y}_{k-1})} \otimes \mathrm{Tr} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y}_{k+1})} \otimes \cdots \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y}_n)}, \tag{2.31}$$

where it is to be understood that the trace mapping on the right-hand-side of this equation acts on $\mathrm{L}(\mathcal{Y}_k)$.

Under the assumption that the classical state sets of $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ are equal to $\Gamma_1, \ldots, \Gamma_n$, respectively, one may define $\sigma = \rho[\mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n]$ explicitly as

$$\begin{aligned} &\sigma\big((a_1, \ldots, a_{k-1}, a_{k+1}, \ldots, a_n), (b_1, \ldots, b_{k-1}, b_{k+1}, \ldots, b_n)\big) \\ &\quad = \sum_{c \in \Gamma_k} \rho\big((a_1, \ldots, a_{k-1}, c, a_{k+1}, \ldots, a_n), (b_1, \ldots, b_{k-1}, c, b_{k+1}, \ldots, b_n)\big) \end{aligned} \tag{2.32}$$

for each choice of $a_j, b_j \in \Gamma_j$ and j ranging over the set $\{1, \ldots, n\} \setminus \{k\}$.

Example 2.7. Let Y and Z be registers, both having the classical state set Σ, let $\mathsf{X} = (\mathsf{Y}, \mathsf{Z})$, and let $u \in \mathcal{X} = \mathcal{Y} \otimes \mathcal{Z}$ be defined as

$$u = \frac{1}{\sqrt{|\Sigma|}} \sum_{a \in \Sigma} e_a \otimes e_a, \tag{2.33}$$

[2] It should be noted that reductions of states are determined in this way, by means of the partial trace, by necessity: no other choice is consistent with the basic notions concerning channels and measurements to be discussed in the sections following this one.

so that

$$uu^* = \frac{1}{|\Sigma|} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b}. \tag{2.34}$$

It holds that

$$\big(uu^*\big)[\mathsf{Y}] = \frac{1}{|\Sigma|} \sum_{a,b \in \Sigma} \mathrm{Tr}(E_{a,b}) E_{a,b} = \frac{1}{|\Sigma|} \mathbb{1}_{\mathcal{Y}}. \tag{2.35}$$

The state uu^* is the canonical example of a *maximally entangled* state of two registers sharing the classical state set Σ.

By applying this definition iteratively, one finds that each state ρ of the register $(\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ uniquely determines the state of

$$\big(\mathsf{Y}_{k_1}, \ldots, \mathsf{Y}_{k_m}\big), \tag{2.36}$$

for $k_1, \ldots, k_m$ being any choice of indices satisfying $1 \leq k_1 < \cdots < k_m \leq n$. The state determined by this process is denoted $\rho[\mathsf{Y}_{k_1}, \ldots, \mathsf{Y}_{k_m}]$ and again is called the reduction of ρ to $(\mathsf{Y}_{k_1}, \ldots, \mathsf{Y}_{k_m})$.

The definition above may be generalized in a natural way so that it allows one to specify the states that result from removing an arbitrary collection of subregisters from a given compound register (assuming that this removal results in a valid register). For the registers described in Example 2.2, for instance, removing the subregister Z_3 from X while it is in the state ρ would leave the resulting register in the state

$$\big(\mathbb{1}_{\mathrm{L}(\mathcal{Y}_1)} \otimes (\mathbb{1}_{\mathrm{L}(\mathcal{Z}_1)} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z}_2)} \otimes \mathrm{Tr}\,)\big)(\rho), \tag{2.37}$$

with the understanding that the trace mapping is defined with respect to $\mathcal{Z}_3$. The pattern represented by this example, in which identity mappings and trace mappings are tensored in accordance with the structure of the register under consideration, is generalized in the most straightforward way to other examples. While it is possible to formalize this definition in complete generality, there is little point in doing so for the purposes of this book: all of the instances of state reductions to be encountered are either cases where the reductions take the form $\rho[\mathsf{Y}_{k_1}, \ldots, \mathsf{Y}_{k_m}]$, as discussed above, or are easily specified explicitly as in the case of the example (2.37) just mentioned.

Purifications of states and operators

In a variety of situations that arise in quantum information theory, wherein a given register X is being considered, it is useful to assume (or simply to

imagine) that X is a subregister of a compound register (X, Y), and to view a given state $\rho \in \mathrm{D}(\mathcal{X})$ of X as having been obtained as a reduction

$$\rho = (uu^*)[\mathsf{X}] = \mathrm{Tr}_{\mathcal{Y}}\bigl(uu^*\bigr) \tag{2.38}$$

of some pure state uu^* of (X, Y). It is natural to ask what the possible states of X are that can arise from a pure state of (X, Y) in this way. This question has a simple answer (to be justified shortly): a state $\rho \in \mathrm{D}(\mathcal{X})$ of X can arise in this way if and only if the rank of ρ does not exceed the number of classical states of the register Y removed from (X, Y) to obtain X.

The following definition is representative of the situation just described. The notion of a *purification* that it defines is used extensively throughout the remainder of the book.

Definition 2.8. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator, and let $u \in \mathcal{X} \otimes \mathcal{Y}$ be a vector. The vector u is a *purification* of P if and only if

$$\mathrm{Tr}_{\mathcal{Y}}\bigl(uu^*\bigr) = P. \tag{2.39}$$

This definition deviates slightly from the setting described above in two respects. One is that the operator P is not required to be a density operator, and the other is that the vector u is taken to be the object that purifies P rather than the operator uu^*. Allowing P to be an arbitrary positive semidefinite operator is a useful generalization that will cause no difficulties in developing the concept of a purification, while referring to u rather than uu^* as the purification of P is simply a matter of convenience based on the specific ways that the notion is most typically used.

It is straightforward to generalize the notion of a purification. One may, for instance, consider the situation in which X is a register that is obtained by removing one or more subregisters from an arbitrary compound register Z. A purification of a given state $\rho \in \mathrm{D}(\mathcal{X})$ in this context would refer to any pure state uu^* of Z whose reduction to X is equal to ρ. In the interest of simplicity, however, it is helpful to restrict one's attention to the specific notion of a purification given by Definition 2.8. All of the interesting aspects of purifications in this restricted setting extend easily and directly to this more general notion of a purification.

Conditions for the existence of purifications

The study of purifications is simplified through the use of the vec mapping defined in Section 1.1.2. Given that the vec mapping is a linear bijection from $\mathrm{L}(\mathcal{Y}, \mathcal{X})$ to $\mathcal{X} \otimes \mathcal{Y}$, every vector $u \in \mathcal{X} \otimes \mathcal{Y}$ may be written as $u = \mathrm{vec}(A)$ for some choice of an operator $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$. By the identity (1.129), it holds that

$$\mathrm{Tr}_{\mathcal{Y}}(uu^*) = \mathrm{Tr}_{\mathcal{Y}}(\mathrm{vec}(A)\,\mathrm{vec}(A)^*) = AA^*, \tag{2.40}$$

establishing an equivalence between the following statements, for a given choice of $P \in \mathrm{Pos}(\mathcal{X})$:

1. There exists a purification $u \in \mathcal{X} \otimes \mathcal{Y}$ of P.
2. There exists an operator $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ such that $P = AA^*$.

The next theorem, whose proof is based on this observation, justifies the answer given above to the question on necessary and sufficient conditions for the existence of a purification of a given operator.

Theorem 2.9. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator. There exists a vector $u \in \mathcal{X} \otimes \mathcal{Y}$ such that $\mathrm{Tr}_{\mathcal{Y}}(uu^*) = P$ if and only if $\dim(\mathcal{Y}) \geq \mathrm{rank}(P)$.*

Proof. As observed above, the existence of a vector $u \in \mathcal{X} \otimes \mathcal{Y}$ for which $\mathrm{Tr}_{\mathcal{Y}}(uu^*) = P$ is equivalent to the existence of an operator $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ satisfying $P = AA^*$. Under the assumption that such an operator A exists, it must hold that $\mathrm{rank}(P) = \mathrm{rank}(A)$, and therefore $\dim(\mathcal{Y}) \geq \mathrm{rank}(P)$.

Conversely, under the assumption $\dim(\mathcal{Y}) \geq \mathrm{rank}(P)$, one may prove the existence of an operator $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ satisfying $P = AA^*$ as follows. Let $r = \mathrm{rank}(P)$ and use the spectral theorem (Corollary 1.4) to write

$$P = \sum_{k=1}^{r} \lambda_k(P)\, x_k x_k^* \tag{2.41}$$

for $\{x_1, \ldots, x_r\} \subset \mathcal{X}$ being an orthonormal set. For $\{y_1, \ldots, y_r\} \subset \mathcal{Y}$ being an arbitrary choice of an orthonormal set in $\mathcal{Y}$, which exists by the assumption $\dim(\mathcal{Y}) \geq \mathrm{rank}(P)$, the operator

$$A = \sum_{k=1}^{r} \sqrt{\lambda_k(P)}\, x_k y_k^* \tag{2.42}$$

satisfies $AA^* = P$. □

Corollary 2.10. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and assume that that $\dim(\mathcal{Y}) \geq \dim(\mathcal{X})$. For every positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X})$, there exists a vector $u \in \mathcal{X} \otimes \mathcal{Y}$ such that $\mathrm{Tr}_{\mathcal{Y}}(uu^*) = P$.*

Unitary equivalence of purifications

Having established a simple condition under which a purification of a given positive semidefinite operator exists, it is natural to consider the possible relationships among different purifications of such an operator. The following theorem establishes a useful relationship between purifications that must always hold.

Theorem 2.11 (Unitary equivalence of purifications)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and suppose that vectors $u, v \in \mathcal{X} \otimes \mathcal{Y}$ satisfy*

$$\mathrm{Tr}_{\mathcal{Y}}(uu^*) = \mathrm{Tr}_{\mathcal{Y}}(vv^*). \tag{2.43}$$

There exists a unitary operator $U \in \mathrm{U}(\mathcal{Y})$ such that $v = (\mathbb{1}_{\mathcal{X}} \otimes U)u$.

Proof. Let $A, B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be the unique operators satisfying $u = \mathrm{vec}(A)$ and $v = \mathrm{vec}(B)$, and let $P \in \mathrm{Pos}(\mathcal{X})$ satisfy

$$\mathrm{Tr}_{\mathcal{Y}}(uu^*) = P = \mathrm{Tr}_{\mathcal{Y}}(vv^*). \tag{2.44}$$

It therefore holds that $AA^* = P = BB^*$. Letting $r = \mathrm{rank}(P)$, it follows that $\mathrm{rank}(A) = r = \mathrm{rank}(B)$.

Now, let $x_1, \ldots, x_r \in \mathcal{X}$ be any orthonormal sequence of eigevectors of P with corresponding eigenvalues $\lambda_1(P), \ldots, \lambda_r(P)$. As $AA^* = P = BB^*$, it is possible (as discussed in Section 1.1.3) to select singular value decompositions

$$A = \sum_{k=1}^{r} \sqrt{\lambda_k(P)}\, x_k y_k^* \quad \text{and} \quad B = \sum_{k=1}^{r} \sqrt{\lambda_k(P)}\, x_k w_k^* \tag{2.45}$$

of A and B, for some choice of orthonormal collections $\{y_1, \ldots, y_r\}$ and $\{w_1, \ldots, w_r\}$ of vectors in $\mathcal{Y}$.

Finally, let $V \in \mathrm{U}(\mathcal{Y})$ be any unitary operator satisfying $Vw_k = y_k$ for every $k \in \{1, \ldots, r\}$. It follows that $AV = B$, and by taking $U = V^{\mathsf{T}}$ one has

$$(\mathbb{1}_{\mathcal{X}} \otimes U)u = \left(\mathbb{1}_{\mathcal{X}} \otimes V^{\mathsf{T}}\right) \mathrm{vec}(A) = \mathrm{vec}(AV) = \mathrm{vec}(B) = v, \tag{2.46}$$

as required. □

2.2 Quantum channels

Quantum channels represent discrete changes in states of registers that are to be considered physically realizable (in an idealized sense). For example, the steps of a quantum computation, or any other processing of quantum information, as well as the effects of errors and noise on quantum registers, are modeled as quantum channels.

2.2.1 Definitions and basic notions concerning channels

In mathematical terms, a quantum channel is a linear map, from one space of square operators to another, that satisfies the two conditions of *complete positivity* and *trace preservation*.

Definition 2.12. A *quantum channel* (or simply a *channel*, for short) is a linear map

$$\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y}) \tag{2.47}$$

(i.e., an element $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$), for some choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, satisfying two properties:

1. Φ is completely positive.
2. Φ is trace-preserving.

The collection of all channels of the form (2.47) is denoted $\mathrm{C}(\mathcal{X}, \mathcal{Y})$, and one writes $\mathrm{C}(\mathcal{X})$ as a shorthand for $\mathrm{C}(\mathcal{X}, \mathcal{X})$.

For a given choice of registers X and Y, one may view that a channel of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is a transformation from X into Y. That is, when such a transformation takes place, it is to be viewed that the register X ceases to exist, with Y being formed in its place. Moreover, the state of Y is obtained by applying the map Φ to the state $\rho \in \mathrm{D}(\mathcal{X})$ of X, yielding $\Phi(\rho) \in \mathrm{D}(\mathcal{Y})$. When it is the case that $\mathsf{X} = \mathsf{Y}$, one may simply view that the state of the register X has been changed according to the mapping Φ.

Example 2.13. Let $\mathcal{X}$ be a complex Euclidean space and let $U \in \mathrm{U}(\mathcal{X})$ be a unitary operator. The map $\Phi \in \mathrm{C}(\mathcal{X})$ defined by

$$\Phi(X) = UXU^* \tag{2.48}$$

for every $X \in \mathrm{L}(\mathcal{X})$ is an example of a channel. Channels of this form are called *unitary channels*. The identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{X})}$ is one example of a unitary channel, obtained by setting $U = \mathbb{1}_{\mathcal{X}}$. Intuitively speaking, this channel represents an ideal communication channel or a perfect component in a quantum computer memory, which causes no change in the state of the register X it acts upon.

Example 2.14. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and let $\sigma \in \mathrm{D}(\mathcal{Y})$ be a density operator. The mapping $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ defined by

$$\Phi(X) = \mathrm{Tr}(X)\sigma \tag{2.49}$$

for every $X \in \mathrm{L}(\mathcal{X})$ is a channel. It holds that $\Phi(\rho) = \sigma$ for every $\rho \in \mathrm{D}(\mathcal{X})$; in effect, the channel Φ represents the action of discarding a given register X, and replacing it with the register Y initialized in the state σ. Channels of this form will be called *replacement channels*.

The channels described in the two previous examples (along with other examples of channels) will be discussed in greater detail in Section 2.2.3. While one may prove directly that these mappings are indeed channels, these facts will follow immediately from more general results to be presented in Section 2.2.2.

Product channels

Suppose that $\mathsf{X}_1, \ldots, \mathsf{X}_n$ and $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ are registers having associated complex Euclidean spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$ and $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$, respectively. A channel

$$\Phi \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n) \tag{2.50}$$

transforming $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ into $(\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ is said to be a *product channel* if and only if

$$\Phi = \Psi_1 \otimes \cdots \otimes \Psi_n \tag{2.51}$$

for some choice of channels $\Psi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$, $\ldots$, $\Psi_n \in \mathrm{C}(\mathcal{X}_n, \mathcal{Y}_n)$. Product channels represent an independent application of a sequence of channels to a sequence of registers, in a similar way to product states representing independence among the states of registers.

An important special case involving independent channels is the situation in which a given channel is performed on one register, while nothing at

all is done to one or more other registers under consideration. (As suggested in Example 2.13, the act of doing nothing at all to a register is equivalent to performing the identity channel on that register.)

Example 2.15. Suppose that X, Y, and Z are registers, and $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is a quantum channel that transforms X into Y. Also suppose that the compound register (X, Z) is in some particular state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$ at some instant, and the channel Φ is applied to X, transforming it into Y. The resulting state of the pair (Y, Z) is then given by

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho) \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z}), \tag{2.52}$$

as one views that the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}$ has independently been applied to the register Z.

Example 2.15 illustrates the importance of the complete positivity requirement on quantum channels. That is, it must hold that $(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\rho)$ is a density operator for every choice of Z and every density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$, which together with the linearity of Φ implies that Φ is completely positive (in addition to being trace-preserving).

State preparations as quantum channels

As stated in Section 2.1.1, a register is *trivial* if its classical state set consists of a single element. The complex Euclidean space associated with a trivial register is therefore one-dimensional: it must take the form $\mathbb{C}^{\{a\}}$ for $\{a\}$ being the singleton classical state set of the register. No generality is lost in associating such a space with the field of complex numbers $\mathbb{C}$, and in making the identification $\mathrm{L}(\mathbb{C}) = \mathbb{C}$, one finds that the scalar 1 is the only possible state for a trivial register. As is to be expected, such a register is therefore completely useless from an information-processing viewpoint; the presence of a trivial register does nothing more than to tensor the scalar 1 to the state of any other registers under consideration.

It is instructive nevertheless to consider the properties of channels that involve trivial registers. Suppose, in particular, that X is a trivial register and Y is arbitrary, and consider a channel of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ that transforms X into Y. It must hold that Φ is given by

$$\Phi(\alpha) = \alpha\rho \tag{2.53}$$

for all $\alpha \in \mathbb{C}$, for some choice of $\rho \in \mathrm{D}(\mathcal{Y})$, as Φ must be linear and it must hold that $\Phi(1)$ is positive semidefinite and has trace equal to one. The channel Φ defined by (2.53) may be viewed as the *preparation* of the quantum state ρ in a new register Y. The trivial register X can be considered as being essentially a placeholder for this preparation, which is to occur at whatever moment the channel Φ is performed. In this way, a state preparation may be seen as the application of this form of channel.

To see that every mapping of the form (2.53) is indeed a channel, for an arbitrary choice of a density operator $\rho \in \mathrm{D}(\mathcal{Y})$, one may check that the conditions of complete positivity and trace preservation hold. The mapping Φ given by (2.53) is obviously trace-preserving whenever $\mathrm{Tr}(\rho) = 1$, and the complete positivity of Φ is implied by the following simple proposition.

Proposition 2.16. *Let $\mathcal{Y}$ be a complex Euclidean space and let $P \in \mathrm{Pos}(\mathcal{Y})$ be a positive semidefinite operator. The mapping $\Phi \in \mathrm{T}(\mathbb{C}, \mathcal{Y})$ defined as $\Phi(\alpha) = \alpha P$ for all $\alpha \in \mathbb{C}$ is completely positive.*

Proof. Let $\mathcal{Z}$ be any complex Euclidean space. The action of the mapping $\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}$ on an operator $Z \in \mathrm{L}(\mathcal{Z}) = \mathrm{L}(\mathbb{C} \otimes \mathcal{Z})$ is given by

$$\bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\bigr)(Z) = P \otimes Z. \tag{2.54}$$

If Z is positive semidefinite, then $P \otimes Z$ is positive semidefinite as well, and therefore Φ is completely positive. □

The trace mapping as a channel

The other situation of a channel involving a trivial register is the one in which a channel Φ transforms an arbitrary register X into a trivial register Y. By identifying the complex Euclidean space $\mathcal{Y}$ with the complex numbers $\mathbb{C}$ as before, one has that the channel Φ must take the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathbb{C})$.

The only mapping of this form that can possibly preserve trace is the trace mapping itself, and so it must hold that

$$\Phi(X) = \mathrm{Tr}(X) \tag{2.55}$$

for all $X \in \mathrm{L}(\mathcal{X})$. To say that a register X has been transformed into a trivial register Y is tantamount to saying that X has been destroyed, discarded, or simply ignored (assuming that the trivial register Y is left unmentioned).

This channel was, in effect, introduced in Section 2.1.3 when reductions of quantum states were defined.

In order to conclude that the trace mapping is indeed a valid channel, it is necessary to verify that it is completely positive. One way to prove this simple fact is to combine the following proposition with Proposition 2.16.

Proposition 2.17. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a positive map. It holds that Φ^* is positive.*

Proof. By the positivity of Φ, it holds that

$$\Phi(P) \in \mathrm{Pos}(\mathcal{Y}) \tag{2.56}$$

for every positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X})$, which is equivalent to the condition that

$$\langle Q, \Phi(P) \rangle \geq 0 \tag{2.57}$$

for all positive semidefinite operators $P \in \mathrm{Pos}(\mathcal{X})$ and $Q \in \mathrm{Pos}(\mathcal{Y})$. It follows that

$$\langle \Phi^*(Q), P \rangle = \langle Q, \Phi(P) \rangle \geq 0 \tag{2.58}$$

for all $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z})$ and $Q \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$, which is equivalent to

$$\Phi^*(Q) \in \mathrm{Pos}(\mathcal{X}) \tag{2.59}$$

for every $Q \in \mathrm{Pos}(\mathcal{Y})$. The mapping Φ^* is therefore positive. □

Remark 2.18. Proposition 2.17 implies, for every completely positive map $\Phi \in \mathrm{CP}(\mathcal{X}, \mathcal{Y})$, that the adjoint mapping Φ^* is also completely positive; for if Φ is completely positive, then $\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}$ is positive for every complex Euclidean space $\mathcal{Z}$, and therefore $(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})^* = \Phi^* \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}$ is also positive.

Corollary 2.19. *The trace mapping $\mathrm{Tr} \in \mathrm{T}(\mathcal{X}, \mathbb{C})$, for any choice of a complex Euclidean space $\mathcal{X}$, is completely positive.*

Proof. The adjoint of the trace is given by $\mathrm{Tr}^*(\alpha) = \alpha \mathbb{1}_{\mathcal{X}}$ for every $\alpha \in \mathbb{C}$. This map is completely positive by Proposition 2.16, therefore the trace map is completely positive by Remark 2.18 to Proposition 2.17. □

2.2.2 Representations and characterizations of channels

Suppose $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is a channel, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. In some situations it may be sufficient to view such a channel abstractly, as a completely positive and trace-preserving linear map of the form $\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$, and nothing more. In other situations, it may be useful to consider a more concrete representation of such a channel.

Four specific representations of channels (and of arbitrary mappings of the form $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$) are discussed in this section. These different representations reveal interesting properties of channels, and will find uses in different situations throughout this book. The simple relationships among the representations generally allow one to convert from one representation into another, and therefore to choose the representation that is best suited to a given situation.

The natural representation

For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and for every linear mapping $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, it is evident that the mapping

$$\operatorname{vec}(X) \mapsto \operatorname{vec}(\Phi(X)) \tag{2.60}$$

is linear, as it can be represented as a composition of linear mappings. There must therefore exist a linear operator $K(\Phi) \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X}, \mathcal{Y} \otimes \mathcal{Y})$ for which

$$K(\Phi) \operatorname{vec}(X) = \operatorname{vec}(\Phi(X)) \tag{2.61}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The operator $K(\Phi)$, which is uniquely determined by the requirement that (2.61) holds for all $X \in \mathrm{L}(\mathcal{X})$, is the *natural representation* of Φ, as it directly represents the action of Φ as a linear mapping (with respect to the operator-vector correspondence).

It may be noted that the mapping $K : \mathrm{T}(\mathcal{X}, \mathcal{Y}) \to \mathrm{L}(\mathcal{X} \otimes \mathcal{X}, \mathcal{Y} \otimes \mathcal{Y})$ is linear:

$$K(\alpha\Phi + \beta\Psi) = \alpha K(\Phi) + \beta K(\Psi) \tag{2.62}$$

for all choices of $\alpha, \beta \in \mathbb{C}$ and $\Phi, \Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$. Moreover, K is a bijection, as the action of a given mapping Φ can be recovered from $K(\Phi)$; for each operator $X \in \mathrm{L}(\mathcal{X})$, one has that $Y = \Phi(X)$ is the unique operator satisfying $\operatorname{vec}(Y) = K(\Phi) \operatorname{vec}(X)$.

The natural representation respects the notion of adjoints, meaning that

$$K(\Phi^*) = (K(\Phi))^* \tag{2.63}$$

for every mappings $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ (with the understanding that K refers to a mapping from $\mathrm{T}(\mathcal{Y}, \mathcal{X})$ to $\mathrm{L}(\mathcal{Y} \otimes \mathcal{Y}, \mathcal{X} \otimes \mathcal{X})$ on the left-hand side of this equation, obtained by reversing the roles of $\mathcal{X}$ and $\mathcal{Y}$ in the definition above).

Despite the fact that the natural representation $K(\Phi)$ of a mapping Φ is a direct representation of the action of Φ as a linear map, this representation is the one of the four representations to be discussed in this section that is the least directly connected to the properties of complete positivity and trace preservation. As such, it will turn out to be the least useful of the four representations from the viewpoint of this book. One explanation for why this is so is that the aspects of a given map Φ that relate to the operator structure of its input and output arguments is not represented by $K(\Phi)$ in a convenient or readily accessible form. The operator-vector correspondence has the effect of ignoring this structure.

The Choi representation

For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, one may define a mapping $J : \mathrm{T}(\mathcal{X}, \mathcal{Y}) \to \mathrm{L}(\mathcal{Y} \otimes \mathcal{X})$ as

$$J(\Phi) = \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)\big(\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*\big) \tag{2.64}$$

for each $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$. Alternatively, under the assumption that $\mathcal{X} = \mathbb{C}^{\Sigma}$, one may write

$$J(\Phi) = \sum_{a,b \in \Sigma} \Phi(E_{a,b}) \otimes E_{a,b}. \tag{2.65}$$

The operator $J(\Phi)$ is called the *Choi representation* (or the *Choi operator*) of Φ.

It is evident from the equation (2.65) that the mapping J is a linear bijection. An alternate way to prove that the mapping J is a bijection is to observe that the action of the mapping Φ can be recovered from the operator $J(\Phi)$ by means of the equation

$$\Phi(X) = \mathrm{Tr}_{\mathcal{X}}\big(J(\Phi)\big(\mathbb{1}_{\mathcal{Y}} \otimes X^{\mathsf{T}}\big)\big). \tag{2.66}$$

There is a close connection between the operator structure of $J(\Phi)$ and the aspects of Φ that relate to the operator structure of its input and output arguments. A central component of this connection is that a given map

Φ is completely positive if and only if $J(\Phi)$ is positive semidefinite (as is established by Theorem 2.22 below).

For a given map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, the rank of the Choi representation $J(\Phi)$ is called the *Choi rank* of Φ.

Kraus representations

For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, an alphabet Σ, and collections

$$\{A_a : a \in \Sigma\} \quad \text{and} \quad \{B_a : a \in \Sigma\} \tag{2.67}$$

of operators drawn from the space $\mathrm{L}(\mathcal{X}, \mathcal{Y})$, one may define a linear map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ as

$$\Phi(X) = \sum_{a \in \Sigma} A_a X B_a^* \tag{2.68}$$

for every $X \in \mathrm{L}(\mathcal{X})$. The expression (2.68) is a *Kraus representation* of the map Φ. It will be established shortly that a Kraus representation exists for every map of the form $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$. Unlike the natural representation and Choi representation, however, Kraus representations are not unique.

Under the assumption that Φ is determined by the above equation (2.68), it holds that

$$\Phi^*(Y) = \sum_{a \in \Sigma} A_a^* Y B_a, \tag{2.69}$$

as follows from a calculation relying on the cyclic property of the trace:

$$\begin{aligned} \left\langle Y, \sum_{a \in \Sigma} A_a X B_a^* \right\rangle &= \sum_{a \in \Sigma} \mathrm{Tr}\bigl(Y^* A_a X B_a^*\bigr) \\ &= \sum_{a \in \Sigma} \mathrm{Tr}\bigl(B_a^* Y^* A_a X\bigr) = \left\langle \sum_{a \in \Sigma} A_a^* Y B_a, X \right\rangle \end{aligned} \tag{2.70}$$

for every $X \in \mathrm{L}(\mathcal{X})$ and $Y \in \mathrm{L}(\mathcal{Y})$.

It is common in the theory of quantum information that one encounters Kraus representations for which $A_a = B_a$ for each $a \in \Sigma$. As is established by Theorem 2.22 below, such representations exist precisely when the map being considered is completely positive.

Stinespring representations

Suppose $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are complex Euclidean spaces and $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ are operators. One may then define a map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ as

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\bigl(AXB^*\bigr) \tag{2.71}$$

for every $X \in \mathrm{L}(\mathcal{X})$. The expression (2.71) is a *Stinespring representation* of the map Φ. Similar to Kraus representations, Stinespring representations always exist for a given map Φ, and are not unique.

If a map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ has a Stinespring representation taking the form (2.71), then it holds that

$$\Phi^*(Y) = A^*(Y \otimes \mathbb{1}_{\mathcal{Z}})B \tag{2.72}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$. This observation follows from a calculation:

$$\begin{aligned} \langle Y, \Phi(X)\rangle &= \langle Y, \mathrm{Tr}_{\mathcal{Z}}(AXB^*)\rangle = \langle Y \otimes \mathbb{1}_{\mathcal{Z}}, AXB^*\rangle \\ &= \mathrm{Tr}\bigl((Y \otimes \mathbb{1}_{\mathcal{Z}})^* AXB^*\bigr) = \mathrm{Tr}\bigl(B^*(Y \otimes \mathbb{1}_{\mathcal{Z}})^* AX\bigr) \\ &= \langle A^*(Y \otimes \mathbb{1}_{\mathcal{Z}})B, X\rangle \end{aligned} \tag{2.73}$$

for every $X \in \mathrm{L}(\mathcal{X})$ and $Y \in \mathrm{L}(\mathcal{Y})$. Expressions of the form (2.72) are also sometimes referred to as Stinespring representations, although the terminology will not be used in this way in this book.

Similar to Kraus representations, it is common in quantum information theory that one encounters Stinespring representations for which $A = B$. Also similar to Kraus representations, such representations exist if and only if Φ is completely positive.

Relationships among the representations

The following proposition relates the four representations discussed above to one another, and (implicitly) shows how any one of the representations may be converted into any another.

Proposition 2.20. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let Σ be an alphabet, let $\{A_a : a \in \Sigma\}, \{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$ be collections of operators indexed by Σ, and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$. The following four statements, which correspond as indicated to the four representations introduced above, are equivalent:*

1. *(Natural representation.) It holds that*

$$K(\Phi) = \sum_{a \in \Sigma} A_a \otimes \overline{B_a}. \tag{2.74}$$

2. *(Choi representation.) It holds that*

$$J(\Phi) = \sum_{a \in \Sigma} \operatorname{vec}(A_a) \operatorname{vec}(B_a)^*. \tag{2.75}$$

3. *(Kraus representations.) It holds that*

$$\Phi(X) = \sum_{a \in \Sigma} A_a X B_a^* \tag{2.76}$$

   for all $X \in \mathrm{L}(\mathcal{X})$.

4. *(Stinespring representations.) For* $\mathcal{Z} = \mathbb{C}^\Sigma$ *and* $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ *defined as*

$$A = \sum_{a \in \Sigma} A_a \otimes e_a \quad \text{and} \quad B = \sum_{a \in \Sigma} B_a \otimes e_a, \tag{2.77}$$

   it holds that

$$\Phi(X) = \operatorname{Tr}_{\mathcal{Z}}\bigl(AXB^*\bigr) \tag{2.78}$$

   for all $X \in \mathrm{L}(\mathcal{X})$.

Proof. The equivalence between statements 3 and 4 is a straightforward calculation. The equivalence between statements 1 and 3 follows from the identity

$$\operatorname{vec}(A_a X B_a^*) = \bigl(A_a \otimes \overline{B_a}\bigr) \operatorname{vec}(X) \tag{2.79}$$

for all choices of $a \in \Sigma$ and $X \in \mathrm{L}(\mathcal{X})$. Finally, the equivalence between statements 2 and 3 follows from the equations

$$\begin{aligned} (A_a \otimes \mathbb{1}_{\mathcal{X}}) \operatorname{vec}(\mathbb{1}_{\mathcal{X}}) &= \operatorname{vec}(A_a), \\ \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^* (B_a^* \otimes \mathbb{1}_{\mathcal{X}}) &= \operatorname{vec}(B_a)^*, \end{aligned} \tag{2.80}$$

which hold for every $a \in \Sigma$. □

Corollary 2.21. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a nonzero linear map, and let $r = \mathrm{rank}(J(\Phi))$ be the Choi rank of Φ. The following two facts hold:*

1. *For Σ being any alphabet with $|\Sigma| = r$, there exists a Kraus representation of Φ having the form*
$$\Phi(X) = \sum_{a \in \Sigma} A_a X B_a^*, \tag{2.81}$$
for some choice of $\{A_a : a \in \Sigma\}, \{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$.
2. *For $\mathcal{Z}$ being any complex Euclidean space with $\dim(\mathcal{Z}) = r$, there exists a Stinespring representation of Φ having the form*
$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\left(AXB^*\right), \tag{2.82}$$
for some choice of operators $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$.

Proof. For Σ being any alphabet with $|\Sigma| = r$, it is possible to write
$$J(\Phi) = \sum_{a \in \Sigma} u_a v_a^* \tag{2.83}$$
for some choice of vectors
$$\{u_a : a \in \Sigma\}, \{v_a : a \in \Sigma\} \subset \mathcal{Y} \otimes \mathcal{X}. \tag{2.84}$$
In particular, one may take $\{u_a \; a \in \Sigma\}$ to be any basis for the subspace $\mathrm{im}(J(\Phi))$, which uniquely determines a collection $\{v_a : a \in \Sigma\}$ for which (2.83) holds. Taking $\{A_a : a \in \Sigma\}$ and $\{B_a : a \in \Sigma\}$ to be operators defined by the equations
$$\mathrm{vec}(A_a) = u_a \quad \text{and} \quad \mathrm{vec}(B_a) = v_a \tag{2.85}$$
for every $a \in \Sigma$, it follows from Proposition 2.20 that (2.81) is a Kraus representation of Φ. Moreover, it holds that (2.82) is a Stinespring representation of Φ for $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ defined as
$$A = \sum_{a \in \Sigma} A_a \otimes e_a \quad \text{and} \quad B = \sum_{a \in \Sigma} B_a \otimes e_a, \tag{2.86}$$
which completes the proof. □

Characterizations of completely positive maps

Characterizations of completely positive maps, based on their Choi, Kraus, and Stinespring representations, will now be presented.

Theorem 2.22. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a nonzero map. The following statements are equivalent:*

1. *Φ is completely positive.*
2. *$\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}$ is positive.*
3. *$J(\Phi) \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$.*
4. *There exists an alphabet Σ and a collection $\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$ for which*
$$\Phi(X) = \sum_{a \in \Sigma} A_a X A_a^* \tag{2.87}$$
for all $X \in \mathrm{L}(\mathcal{X})$.
5. *Statement 4 holds for an alphabet Σ satisfying $|\Sigma| = \mathrm{rank}(J(\Phi))$.*
6. *There exists a complex Euclidean space $\mathcal{Z}$ and an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ such that*
$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\left(AXA^*\right) \tag{2.88}$$
for all $X \in \mathrm{L}(\mathcal{X})$.
7. *Statement 6 holds for $\mathcal{Z}$ having dimension equal to $\mathrm{rank}(J(\Phi))$.*

Proof. The theorem will be proved by establishing implications among the seven statements that are sufficient to imply their equivalence. The implications that will be proved are summarized as follows:

$$\begin{gathered}(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (5) \Rightarrow (4) \Rightarrow (1) \\ (5) \Rightarrow (7) \Rightarrow (6) \Rightarrow (1)\end{gathered}$$

Note that some of these implications are immediate: statement 1 implies statement 2 by the definition of complete positivity, statement 5 trivially implies statement 4, statement 7 trivially implies statement 6, and statement 5 implies statement 7 by Proposition 2.20.

Assume $\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}$ is positive. Given that

$$\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{X}) \tag{2.89}$$

and

$$J(\Phi) = (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(\operatorname{vec}(\mathbb{1}_{\mathcal{X}}) \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^*), \tag{2.90}$$

it follows that $J(\Phi) \in \operatorname{Pos}(\mathcal{Y} \otimes \mathcal{X})$. Statement 2 therefore implies statement 3.

Next, assume $J(\Phi) \in \operatorname{Pos}(\mathcal{Y} \otimes \mathcal{X})$. It follows by the spectral theorem (Corollary 1.4), together with the fact that every eigenvalue of a positive semidefinite operator is nonnegative, that one may write

$$J(\Phi) = \sum_{a \in \Sigma} u_a u_a^*, \tag{2.91}$$

for some choice of an alphabet Σ with $|\Sigma| = \operatorname{rank}(J(\Phi))$ and a collection

$$\{u_a : a \in \Sigma\} \subset \mathcal{Y} \otimes \mathcal{X} \tag{2.92}$$

of vectors. Taking $A_a \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ to be the operator defined by the equation $\operatorname{vec}(A_a) = u_a$ for each $a \in \Sigma$, one has that

$$J(\Phi) = \sum_{a \in \Sigma} \operatorname{vec}(A_a) \operatorname{vec}(A_a)^*. \tag{2.93}$$

The equation (2.87) therefore holds for every $X \in \mathrm{L}(\mathcal{X})$ by Proposition 2.20, which establishes that statement 3 implies statement 5.

Now suppose (2.87) holds for every $X \in \mathrm{L}(\mathcal{X})$, for some alphabet Σ and a collection

$$\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y}) \tag{2.94}$$

of operators. For a complex Euclidean space $\mathcal{W}$ and a positive semidefinite operator $P \in \operatorname{Pos}(\mathcal{X} \otimes \mathcal{W})$, it is evident that

$$(A_a \otimes \mathbb{1}_{\mathcal{W}}) P (A_a \otimes \mathbb{1}_{\mathcal{W}})^* \in \operatorname{Pos}(\mathcal{Y} \otimes \mathcal{W}) \tag{2.95}$$

for each $a \in \Sigma$, and therefore

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(P) \in \operatorname{Pos}(\mathcal{Y} \otimes \mathcal{W}) \tag{2.96}$$

by the convexity of $\operatorname{Pos}(\mathcal{Y} \otimes \mathcal{W})$. It follows that Φ is completely positive, so statement 4 implies statement 1.

Finally, suppose (2.88) holds for every $X \in \mathrm{L}(\mathcal{X})$, for some complex Euclidean space $\mathcal{Z}$ and an operator $A \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$. For any complex Euclidean space $\mathcal{W}$ and any positive semidefinite operator $P \in \operatorname{Pos}(\mathcal{X} \otimes \mathcal{W})$, it is again evident that

$$(A \otimes \mathbb{1}_{\mathcal{W}}) P (A \otimes \mathbb{1}_{\mathcal{W}})^* \in \operatorname{Pos}(\mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{W}), \tag{2.97}$$

and therefore

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(P) = \mathrm{Tr}_{\mathcal{Z}}\big((A \otimes \mathbb{1}_{\mathcal{W}})P(A \otimes \mathbb{1}_{\mathcal{W}})^*\big) \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{W}) \tag{2.98}$$

follows by the complete positivity of the trace (Corollary 2.19). The map Φ is therefore completely positive, so statement 6 implies statement 1, which completes the proof. □

One consequence of this theorem is the following corollary, which relates Kraus representations of a given completely positive map.

Corollary 2.23. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let Σ be an alphabet, and suppose $\{A_a : a \in \Sigma\}$, $\{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$ are collections of operators for which*

$$\sum_{a\in\Sigma} A_a X A_a^* = \sum_{a\in\Sigma} B_a X B_a^* \tag{2.99}$$

for all $X \in \mathrm{L}(\mathcal{X})$. There exists a unitary operator $U \in \mathrm{U}\big(\mathbb{C}^\Sigma\big)$ such that

$$B_a = \sum_{b\in\Sigma} U(a,b) A_b \tag{2.100}$$

for all $a \in \Sigma$.

Proof. The maps

$$X \mapsto \sum_{a\in\Sigma} A_a X A_a^* \quad \text{and} \quad X \mapsto \sum_{a\in\Sigma} B_a X B_a^* \tag{2.101}$$

agree for all $X \in \mathrm{L}(\mathcal{X})$, and therefore their Choi representations must be equal:

$$\sum_{a\in\Sigma} \mathrm{vec}(A_a)\,\mathrm{vec}(A_a)^* = \sum_{a\in\Sigma} \mathrm{vec}(B_a)\,\mathrm{vec}(B_a)^*. \tag{2.102}$$

Let $\mathcal{Z} = \mathbb{C}^\Sigma$ and define vectors $u, v \in \mathcal{Y} \otimes \mathcal{X} \otimes \mathcal{Z}$ as

$$u = \sum_{a\in\Sigma} \mathrm{vec}(A_a) \otimes e_a \quad \text{and} \quad v = \sum_{a\in\Sigma} \mathrm{vec}(B_a) \otimes e_a, \tag{2.103}$$

so that

$$\begin{aligned} \mathrm{Tr}_{\mathcal{Z}}(uu^*) &= \sum_{a\in\Sigma} \mathrm{vec}(A_a)\,\mathrm{vec}(A_a)^* \\ &= \sum_{a\in\Sigma} \mathrm{vec}(B_a)\,\mathrm{vec}(B_a)^* = \mathrm{Tr}_{\mathcal{Z}}(vv^*). \end{aligned} \tag{2.104}$$

By the unitary equivalence of purifications (Theorem 2.11), there must exist a unitary operator $U \in \mathrm{U}(\mathcal{Z})$ such that

$$v = (\mathbb{1}_{\mathcal{Y}\otimes\mathcal{X}} \otimes U)u. \tag{2.105}$$

Thus, for each $a \in \Sigma$ it holds that

$$\operatorname{vec}(B_a) = (\mathbb{1}_{\mathcal{Y}\otimes\mathcal{X}} \otimes e_a^*)v = (\mathbb{1}_{\mathcal{Y}\otimes\mathcal{X}} \otimes e_a^* U)u = \sum_{b\in\Sigma} U(a,b)\operatorname{vec}(A_b), \tag{2.106}$$

which is equivalent to (2.100). □

Along similar lines to the previous corollary is the following one, which concerns Stinespring representations rather than Kraus representations. As the proof reveals, the two corollaries are essentially equivalent.

Corollary 2.24. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces and let operators $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y}\otimes\mathcal{Z})$ satisfy the equation*

$$\operatorname{Tr}_{\mathcal{Z}}\bigl(AXA^*\bigr) = \operatorname{Tr}_{\mathcal{Z}}\bigl(BXB^*\bigr) \tag{2.107}$$

for every $X \in \mathrm{L}(\mathcal{X})$. There exists a unitary operator $U \in \mathrm{U}(\mathcal{Z})$ such that

$$B = (\mathbb{1}_{\mathcal{Y}} \otimes U)A. \tag{2.108}$$

Proof. Let Σ be the alphabet for which $\mathcal{Z} = \mathbb{C}^{\Sigma}$, and define two collections $\{A_a : a \in \Sigma\}$, $\{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$ of operators as

$$A_a = (\mathbb{1}_{\mathcal{Y}} \otimes e_a^*)A \quad \text{and} \quad B_a = (\mathbb{1}_{\mathcal{Y}} \otimes e_a^*)B, \tag{2.109}$$

for each $a \in \Sigma$, so that

$$A = \sum_{a\in\Sigma} A_a \otimes e_a \quad \text{and} \quad B = \sum_{a\in\Sigma} B_a \otimes e_a. \tag{2.110}$$

The equation (2.107) is equivalent to (2.99) in Corollary 2.23. It follows from that corollary that there exists a unitary operator $U \in \mathrm{U}(\mathcal{Z})$ such that (2.100) holds, which is equivalent to $B = (\mathbb{1}_{\mathcal{Y}} \otimes U)A$. □

A map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is said to be *Hermiticity preserving* if it holds that $\Phi(H) \in \mathrm{Herm}(\mathcal{Y})$ for all $H \in \mathrm{Herm}(\mathcal{X})$. The following theorem, which provides four alternative characterizations of this class of maps, is proved through the use of Theorem 2.22.

Theorem 2.25. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map. The following statements are equivalent:*

1. *Φ is Hermiticity preserving.*
2. *It holds that $(\Phi(X))^* = \Phi(X^*)$ for every $X \in \mathrm{L}(\mathcal{X})$.*
3. *It holds that $J(\Phi) \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})$.*
4. *It holds that $\Phi = \Phi_0 - \Phi_1$ for completely positive maps $\Phi_0, \Phi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Y})$.*
5. *It holds that $\Phi = \Phi_0 - \Phi_1$ for positive maps $\Phi_0, \Phi_1 \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$.*

Proof. Assume first that Φ is Hermiticity preserving. For an arbitrary operator $X \in \mathrm{L}(\mathcal{X})$, one may write $X = H + iK$ for $H, K \in \mathrm{Herm}(\mathcal{X})$ being defined as

$$H = \frac{X + X^*}{2} \quad \text{and} \quad K = \frac{X - X^*}{2i}. \tag{2.111}$$

As $\Phi(H)$ and $\Phi(K)$ are both Hermitian and Φ is linear, it follows that

$$\begin{aligned} (\Phi(X))^* &= (\Phi(H) + i\Phi(K))^* \\ &= \Phi(H) - i\Phi(K) = \Phi(H - iK) = \Phi(X^*). \end{aligned} \tag{2.112}$$

Statement 1 therefore implies statement 2.

Next, assume statement 2 holds, and let Σ be the alphabet for which $\mathcal{X} = \mathbb{C}^\Sigma$. One then has that

$$\begin{aligned} J(\Phi)^* = \sum_{a,b \in \Sigma} \Phi(E_{a,b})^* \otimes E_{a,b}^* &= \sum_{a,b \in \Sigma} \Phi(E_{a,b}^*) \otimes E_{a,b}^* \\ &= \sum_{a,b \in \Sigma} \Phi(E_{b,a}) \otimes E_{b,a} = J(\Phi). \end{aligned} \tag{2.113}$$

It follows that $J(\Phi)$ is Hermitian, and therefore statement 3 holds.

Now assume statement 3 holds. Let $J(\Phi) = P_0 - P_1$ be the Jordan–Hahn decomposition of $J(\Phi)$, and let $\Phi_0, \Phi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Y})$ be the maps for which $J(\Phi_0) = P_0$ and $J(\Phi_1) = P_1$. Because P_0 and P_1 are positive semidefinite, it follows from Theorem 2.22 that Φ_0 and Φ_1 are completely positive maps. By the linearity of the mapping J associated with the Choi representation, it holds that $J(\Phi) = J(\Phi_0 - \Phi_1)$, and therefore $\Phi = \Phi_0 - \Phi_1$, implying that statement 4 holds.

Statement 4 trivially implies statement 5.

Finally, assume statement 5 holds. Let $H \in \mathrm{Herm}(\mathcal{X})$ be a Hermitian operator, and let $H = P_0 - P_1$, for $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$, be the Jordan–Hahn

decomposition of H. It holds that $\Phi_a(P_b) \in \mathrm{Pos}(\mathcal{Y})$, for all $a, b \in \{0,1\}$, by the positivity of Φ_0 and Φ_1. Therefore, one has that

$$\Phi(H) = \big(\Phi_0(P_0) + \Phi_1(P_1)\big) - \big(\Phi_0(P_1) + \Phi_1(P_0)\big) \tag{2.114}$$

is the difference between two positive semidefinite operators, and is therefore Hermitian. Thus, statement 1 holds.

As the implications (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (1) among the statements have been established, the theorem is proved. □

Characterizations of trace-preserving maps

The next theorem provides multiple characterizations of the class of trace-preserving maps, presented in a style similar to Theorem 2.22.

Theorem 2.26. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map. The following statements are equivalent:*

1. *Φ is trace-preserving.*
2. *Φ^* is unital.*
3. *$\mathrm{Tr}_{\mathcal{Y}}\big(J(\Phi)\big) = \mathbb{1}_{\mathcal{X}}$.*
4. *There exists a Kraus representation*

$$\Phi(X) = \sum_{a \in \Sigma} A_a X B_a^* \tag{2.115}$$

of Φ for which the operators $\{A_a : a \in \Sigma\}$, $\{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$ satisfy

$$\sum_{a \in \Sigma} A_a^* B_a = \mathbb{1}_{\mathcal{X}}. \tag{2.116}$$

5. *For all Kraus representations of Φ having the form (2.115), the collections of operators $\{A_a : a \in \Sigma\}$ and $\{B_a : a \in \Sigma\}$ satisfy the equation (2.116).*
6. *There exists a Stinespring representation*

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(AXB^*\big) \tag{2.117}$$

of Φ for which the operators $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ satisfy $A^*B = \mathbb{1}_{\mathcal{X}}$.

7. *For all Stinespring representations of Φ of the form (2.117), the operators A and B satisfy $A^*B = \mathbb{1}_{\mathcal{X}}$.*

Proof. Under the assumption that Φ is trace-preserving, it holds that

$$\langle \mathbb{1}_{\mathcal{X}}, X \rangle = \mathrm{Tr}(X) = \mathrm{Tr}(\Phi(X)) = \langle \mathbb{1}_{\mathcal{Y}}, \Phi(X) \rangle = \langle \Phi^*(\mathbb{1}_{\mathcal{Y}}), X \rangle, \tag{2.118}$$

so that

$$\langle \mathbb{1}_{\mathcal{X}} - \Phi^*(\mathbb{1}_{\mathcal{Y}}), X \rangle = 0 \tag{2.119}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It follows that $\Phi^*(\mathbb{1}_{\mathcal{Y}}) = \mathbb{1}_{\mathcal{X}}$, and therefore Φ^* is unital. Along similar lines, the assumption that Φ^* is unital implies

$$\mathrm{Tr}(\Phi(X)) = \langle \mathbb{1}_{\mathcal{Y}}, \Phi(X) \rangle = \langle \Phi^*(\mathbb{1}_{\mathcal{Y}}), X \rangle = \langle \mathbb{1}_{\mathcal{X}}, X \rangle = \mathrm{Tr}(X) \tag{2.120}$$

for every $X \in \mathrm{L}(\mathcal{X})$, and therefore Φ is trace-preserving. The equivalence of statements 1 and 2 has been established.

Next, suppose

$$\Phi(X) = \sum_{a \in \Sigma} A_a X B_a^* \tag{2.121}$$

is a Kraus representation of Φ. It holds that

$$\Phi^*(Y) = \sum_{a \in \Sigma} A_a^* Y B_a \tag{2.122}$$

for every $Y \in \mathrm{L}(\mathcal{Y})$, and in particular it holds that

$$\Phi^*(\mathbb{1}_{\mathcal{Y}}) = \sum_{a \in \Sigma} A_a^* B_a. \tag{2.123}$$

Thus, if Φ^* is unital, then

$$\sum_{a \in \Sigma} A_a^* B_a = \mathbb{1}_{\mathcal{X}}, \tag{2.124}$$

and so it has been proved that statement 2 implies statement 5. On the other hand, if (2.124) holds, then it follows that $\Phi^*(\mathbb{1}_{\mathcal{Y}}) = \mathbb{1}_{\mathcal{X}}$, so that Φ^* is unital. Therefore, statement 4 implies statement 2. As statement 5 trivially implies statement 4, the equivalence of statements 2, 4, and 5 has been established.

Now assume $\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(AXB^*\big)$ is a Stinespring representation of Φ. It follows that

$$\Phi^*(Y) = A^*(Y \otimes \mathbb{1}_{\mathcal{Z}})B \tag{2.125}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$, and in particular $\Phi^*(\mathbb{1}_{\mathcal{Y}}) = A^*B$. The equivalence of statements 2, 6, and 7 follows by the same reasoning as for the case of statements 2, 4, and 5.

Finally, let Γ be the alphabet for which $\mathcal{X} = \mathbb{C}^{\Gamma}$, and consider the operator

$$\mathrm{Tr}_{\mathcal{Y}}(J(\Phi)) = \sum_{a,b\in\Gamma} \mathrm{Tr}(\Phi(E_{a,b}))E_{a,b}. \tag{2.126}$$

If Φ is trace-preserving, then it follows that

$$\mathrm{Tr}(\Phi(E_{a,b})) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b, \end{cases} \tag{2.127}$$

and therefore

$$\mathrm{Tr}_{\mathcal{Y}}(J(\Phi)) = \sum_{a\in\Gamma} E_{a,a} = \mathbb{1}_{\mathcal{X}}. \tag{2.128}$$

Conversely, if $\mathrm{Tr}_{\mathcal{Y}}(J(\Phi)) = \mathbb{1}_{\mathcal{X}}$, then a consideration of the expression (2.126) reveals that (2.127) must hold. The map Φ is therefore trace-preserving by linearity and the fact that $\{E_{a,b} : a, b \in \Gamma\}$ is a basis of $\mathrm{L}(\mathcal{X})$. Statements 1 and 3 are therefore equivalent, which completes the proof. □

Characterizations of channels

Theorems 2.22 and 2.26 can be combined, providing characterizations of channels based on their Choi, Kraus, and Stinespring representations.

Corollary 2.27. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map. The following statements are equivalent:*

1. *Φ is a channel.*
2. *$J(\Phi) \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$ and $\mathrm{Tr}_{\mathcal{Y}}(J(\Phi)) = \mathbb{1}_{\mathcal{X}}$.*
3. *There exists an alphabet Σ and a collection $\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Y})$ satisfying*

   $$\sum_{a\in\Sigma} A_a^* A_a = \mathbb{1}_{\mathcal{X}} \quad \text{and} \quad \Phi(X) = \sum_{a\in\Sigma} A_a X A_a^* \tag{2.129}$$

   for all $X \in \mathrm{L}(\mathcal{X})$.
4. *Statement 3 holds for $|\Sigma| = \mathrm{rank}(J(\Phi))$.*
5. *There exists a complex Euclidean space $\mathcal{Z}$ and an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$, such that*

   $$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\left(AXA^*\right) \tag{2.130}$$

   for all $X \in \mathrm{L}(\mathcal{X})$.
6. *Statement 5 holds under the additional requirement $\dim(\mathcal{Z}) = \mathrm{rank}(J(\Phi))$.*

For every choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, one has that the set of channels $\mathrm{C}(\mathcal{X},\mathcal{Y})$ is compact and convex. One way to prove this fact makes use of the previous corollary.

Proposition 2.28. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces. The set $\mathrm{C}(\mathcal{X},\mathcal{Y})$ is compact and convex.*

Proof. The map $J : \mathrm{T}(\mathcal{X},\mathcal{Y}) \to \mathrm{L}(\mathcal{Y}\otimes\mathcal{X})$ defining the Choi representation is linear and invertible. By Corollary 2.27, one has $J^{-1}(\mathcal{A}) = \mathrm{C}(\mathcal{X},\mathcal{Y})$ for $\mathcal{A}$ being defined as

$$\mathcal{A} = \big\{X \in \mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X}) \,:\, \mathrm{Tr}_{\mathcal{Y}}(X) = \mathbb{1}_{\mathcal{X}}\big\}. \tag{2.131}$$

It therefore suffices to prove that $\mathcal{A}$ is compact and convex. It is evident that $\mathcal{A}$ is closed and convex, as it is the intersection of the positive semidefinite cone $\mathrm{Pos}(\mathcal{Y}\otimes\mathcal{X})$ with the affine subspace

$$\big\{X \in \mathrm{L}(\mathcal{Y}\otimes\mathcal{X}) \,:\, \mathrm{Tr}_{\mathcal{Y}}(X) = \mathbb{1}_{\mathcal{X}}\big\}, \tag{2.132}$$

both of which are closed and convex. To complete the proof, it suffices to prove that $\mathcal{A}$ is bounded. For every $X \in \mathcal{A}$, one has

$$\|X\|_1 = \mathrm{Tr}(X) = \mathrm{Tr}\big(\mathrm{Tr}_{\mathcal{Y}}(X)\big) = \mathrm{Tr}\big(\mathbb{1}_{\mathcal{X}}\big) = \dim(\mathcal{X}), \tag{2.133}$$

and therefore $\mathcal{A}$ is bounded, as required. □

Corollary 2.27 will be used frequently throughout this book, sometimes implicitly. The next proposition, which builds on the unitary equivalence of purifications (Theorem 2.11) to relate a given purification of a positive semidefinite operator to any extension of that operator, is one example of an application of this corollary.

Proposition 2.29. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, and suppose that $u \in \mathcal{X}\otimes\mathcal{Y}$ and $P \in \mathrm{Pos}(\mathcal{X}\otimes\mathcal{Z})$ satisfy*

$$\mathrm{Tr}_{\mathcal{Y}}\big(uu^*\big) = \mathrm{Tr}_{\mathcal{Z}}(P). \tag{2.134}$$

There exists a channel $\Phi \in \mathrm{C}(\mathcal{Y},\mathcal{Z})$ such that

$$\big(\mathbb{1}_{\mathrm{L}(\mathcal{X})}\otimes\Phi\big)(uu^*) = P. \tag{2.135}$$

Proof. Let $\mathcal{W}$ be a complex Euclidean space satisfying

$$\dim(\mathcal{W}) = \max\{\dim(\mathcal{Y}), \operatorname{rank}(P)\}, \tag{2.136}$$

and let $A \in \mathrm{U}(\mathcal{Y}, \mathcal{W})$ be any isometry. Also let $v \in \mathcal{X} \otimes \mathcal{Z} \otimes \mathcal{W}$ satisfy $\operatorname{Tr}_{\mathcal{W}}(vv^*) = P$. It holds that

$$\begin{aligned} &\operatorname{Tr}_{\mathcal{Z}\otimes\mathcal{W}}\bigl((\mathbb{1}_{\mathcal{X}} \otimes A)uu^*(\mathbb{1}_{\mathcal{X}} \otimes A)^*\bigr) \\ &\qquad = \operatorname{Tr}_{\mathcal{Y}}(uu^*) = \operatorname{Tr}_{\mathcal{Z}}(P) = \operatorname{Tr}_{\mathcal{Z}\otimes\mathcal{W}}(vv^*). \end{aligned} \tag{2.137}$$

By Theorem 2.11 there must exist a unitary operator $U \in \mathrm{U}(\mathcal{Z} \otimes \mathcal{W})$ such that

$$(\mathbb{1}_{\mathcal{X}} \otimes UA)u = v. \tag{2.138}$$

Define $\Phi \in \mathrm{T}(\mathcal{Y}, \mathcal{Z})$ as

$$\Phi(Y) = \operatorname{Tr}_{\mathcal{W}}\bigl((UA)Y(UA)^*\bigr) \tag{2.139}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$. By Corollary 2.27, one has that Φ is a channel. It holds that

$$\begin{aligned} \bigl(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi\bigr)(uu^*) &= \operatorname{Tr}_{\mathcal{W}}\bigl((\mathbb{1}_{\mathcal{X}} \otimes UA)uu^*(\mathbb{1}_{\mathcal{X}} \otimes UA)^*\bigr) \\ &= \operatorname{Tr}_{\mathcal{W}}(vv^*) = P, \end{aligned} \tag{2.140}$$

as required. □

2.2.3 Examples of channels and other mappings

This section describes examples of channels, and other maps, along with their specifications according to the four types of representations discussed above. Many other examples and general classifications of channels and maps will be encountered throughout the book.

Isometric and unitary channels

Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $A, B \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ be operators, and consider the map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ defined by

$$\Phi(X) = AXB^* \tag{2.141}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

In the case that $A = B$, and assuming in addition that this operator is a linear isometry from $\mathcal{X}$ to $\mathcal{Y}$, it follows from Corollary 2.27 that Φ is a channel. Such a channel is said to be an *isometric channel*. If $\mathcal{Y} = \mathcal{X}$ and $A = B$ is a unitary operator, Φ is said to be a *unitary channel*. Unitary channels, and convex combinations of unitary channels, are discussed in greater detail in Chapter 4.

The natural representation of the mapping Φ defined by (2.141) is

$$K(\Phi) = A \otimes \overline{B} \tag{2.142}$$

and the Choi representation of Φ is

$$J(\Phi) = \operatorname{vec}(A) \operatorname{vec}(B)^*. \tag{2.143}$$

The expression (2.141) is a Kraus representation of Φ, and may also be regarded as a trivial example of a Stinespring representation if one takes $\mathcal{Z} = \mathbb{C}$ and observes that the trace acts as the identity mapping on $\mathbb{C}$.

The identity mapping $\mathbb{1}_{\mathrm{L}(\mathcal{X})}$ is a simple example of a unitary channel. The natural representation of this channel is the identity operator $\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{X}}$, while its Choi representation is given by the rank-one operator $\operatorname{vec}(\mathbb{1}_{\mathcal{X}}) \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^*$.

Replacement channels and the completely depolarizing channel

Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $A \in \mathrm{L}(\mathcal{X})$ and $B \in \mathrm{L}(\mathcal{Y})$ be operators, and consider the map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ defined as

$$\Phi(X) = \langle A, X \rangle \, B \tag{2.144}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The natural representation of Φ is

$$K(\Phi) = \operatorname{vec}(B) \operatorname{vec}(A)^*, \tag{2.145}$$

and the Choi representation of Φ is

$$J(\Phi) = B \otimes \overline{A}. \tag{2.146}$$

Kraus and Stinespring representations of Φ may also be constructed, although they are not necessarily enlightening in this particular case. One way to obtain a Kraus representation of Φ is to first write

$$A = \sum_{a \in \Sigma} u_a x_a^* \quad \text{and} \quad B = \sum_{b \in \Gamma} v_b y_b^*, \tag{2.147}$$

for some choice of alphabets Σ and Γ and four sets of vectors:

$$\begin{aligned} \{u_a : a \in \Sigma\}, \{x_a : a \in \Sigma\} \subset \mathcal{X}, \\ \{v_b : b \in \Gamma\}, \{y_b : b \in \Gamma\} \subset \mathcal{Y}. \end{aligned} \tag{2.148}$$

It then follows that one Kraus representation of Φ is given by

$$\Phi(X) = \sum_{(a,b)\in\Sigma\times\Gamma} C_{a,b} X D_{a,b}^* \tag{2.149}$$

where $C_{a,b} = v_b u_a^*$ and $D_{a,b} = y_b x_a^*$ for each $a \in \Sigma$ and $b \in \Gamma$, and one Stinespring representation is given by

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(CXD^*), \tag{2.150}$$

where

$$C = \sum_{(a,b)\in\Sigma\times\Gamma} C_{a,b} \otimes e_{(a,b)}, \quad D = \sum_{(a,b)\in\Sigma\times\Gamma} D_{a,b} \otimes e_{(a,b)}, \tag{2.151}$$

and $\mathcal{Z} = \mathbb{C}^{\Sigma\times\Gamma}$.

If A and B are positive semidefinite operators and the map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is defined by (2.144) for all $X \in \mathrm{L}(\mathcal{X})$, then $J(\Phi) = B \otimes \overline{A}$ is positive semidefinite, and therefore Φ is completely positive by Theorem 2.22. In the case that $A = \mathbb{1}_{\mathcal{X}}$ and $B = \sigma$ for some density operator $\sigma \in \mathrm{D}(\mathcal{Y})$, the map Φ is also trace-preserving, and is therefore a channel. As was indicated previously, such a channel is a *replacement channel*—it effectively discards its input, replacing it with the state σ.

The *completely depolarizing channel* $\Omega \in \mathrm{C}(\mathcal{X})$ is an important example of a replacement channel. This channel is defined as

$$\Omega(X) = \mathrm{Tr}(X)\omega \tag{2.152}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for

$$\omega = \frac{\mathbb{1}_{\mathcal{X}}}{\dim(\mathcal{X})} \tag{2.153}$$

denoting the completely mixed state defined with respect to the space $\mathcal{X}$. Equivalently, Ω is the unique channel transforming every density operator into this completely mixed state: $\Omega(\rho) = \omega$ for all $\rho \in \mathrm{D}(\mathcal{X})$. From the

equations (2.145) and (2.146), one has that the natural representation of the completely depolarizing channel $\Omega \in \mathrm{C}(\mathcal{X})$ is

$$K(\Omega) = \frac{\operatorname{vec}(\mathbb{1}_{\mathcal{X}}) \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^*}{\dim(\mathcal{X})}, \tag{2.154}$$

while the Choi representation of this channel is

$$J(\Omega) = \frac{\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{X}}}{\dim(\mathcal{X})}. \tag{2.155}$$

The transpose mapping

Let Σ be an alphabet, let $\mathcal{X} = \mathbb{C}^{\Sigma}$, and let $\mathrm{T} \in \mathrm{T}(\mathcal{X})$ denote the transpose map, defined as

$$\mathrm{T}(X) = X^{\mathsf{T}} \tag{2.156}$$

for all $X \in \mathrm{L}(\mathcal{X})$. This mapping will play an important role in Chapter 6, due to its connections to properties of entangled states.

The natural representation $K(\mathrm{T})$ of T must, by definition, satisfy

$$K(\mathrm{T}) \operatorname{vec}(X) = \operatorname{vec}(X^{\mathsf{T}}) \tag{2.157}$$

for all $X \in \mathrm{L}(\mathcal{X})$. By considering those operators of the form $X = uv^{\mathsf{T}}$ for vectors $u, v \in \mathcal{X}$, one finds that

$$K(\mathrm{T})(u \otimes v) = v \otimes u. \tag{2.158}$$

It follows that $K(\mathrm{T}) = W$, for $W \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$ being the *swap operator*, which is defined by the action $W(u \otimes v) = v \otimes u$ for all vectors $u, v \in \mathcal{X}$.

The Choi representation of T is also equal to the swap operator, as

$$J(\mathrm{T}) = \sum_{a,b \in \Sigma} E_{b,a} \otimes E_{a,b} = W. \tag{2.159}$$

Under the assumption that $|\Sigma| \geq 2$, it therefore follows from Theorem 2.22 that T is not a completely positive map, as W is not a positive semidefinite operator in this case.

One example of a Kraus representation of T is

$$\mathrm{T}(X) = \sum_{a,b \in \Sigma} E_{a,b} X E_{b,a}^* \tag{2.160}$$

for all $X \in \mathrm{L}(\mathcal{X})$, from which it follows that $\mathrm{T}(X) = \operatorname{Tr}_{\mathcal{Z}}(AXB^*)$ is a Stinespring representation of T for $\mathcal{Z} = \mathbb{C}^{\Sigma \times \Sigma}$,

$$A = \sum_{a,b \in \Sigma} E_{a,b} \otimes e_{(a,b)}, \quad \text{and} \quad B = \sum_{a,b \in \Sigma} E_{b,a} \otimes e_{(a,b)}. \tag{2.161}$$

The completely dephasing channel

Let Σ be an alphabet and let $\mathcal{X} = \mathbb{C}^{\Sigma}$. The map $\Delta \in \mathrm{T}(\mathcal{X})$ defined as

$$\Delta(X) = \sum_{a \in \Sigma} X(a,a) E_{a,a} \tag{2.162}$$

for every $X \in \mathrm{L}(\mathcal{X})$ is an example of a channel known as the *completely dephasing channel*. This channel has the effect of replacing every off-diagonal entry of a given operator $X \in \mathrm{L}(\mathcal{X})$ by 0 and leaving the diagonal entries unchanged.

Given the association of diagonal density operators with classical probabilistic states, as discussed in Section 2.1.2, one may view the channel Δ as an ideal classical channel: it acts as the identity mapping on every diagonal density operator, so that it effectively transmits classical probabilistic states without error, while all other states are mapped to the probabilistic states given by their diagonal entries.

The natural representation of Δ must satisfy the equation

$$K(\Delta)\operatorname{vec}(E_{a,b}) = \begin{cases} \operatorname{vec}(E_{a,b}) & \text{if } a = b \\ 0 & \text{if } a \neq b, \end{cases} \tag{2.163}$$

which is equivalent to

$$K(\Delta)(e_a \otimes e_b) = \begin{cases} e_a \otimes e_b & \text{if } a = b \\ 0 & \text{if } a \neq b, \end{cases} \tag{2.164}$$

for every $a, b \in \Sigma$. It follows that

$$K(\Delta) = \sum_{a \in \Sigma} E_{a,a} \otimes E_{a,a}. \tag{2.165}$$

Similar to the transpose mapping, the Choi representation of Δ happens to coincide with its natural representation, as the calculation

$$J(\Delta) = \sum_{a,b \in \Sigma} \Delta(E_{a,b}) \otimes E_{a,b} = \sum_{a \in \Sigma} E_{a,a} \otimes E_{a,a} \tag{2.166}$$

reveals. It is evident from this expression, together with Corollary 2.27, that Δ is indeed a channel.

One example of a Kraus representation of Δ is

$$\Delta(X) = \sum_{a \in \Sigma} E_{a,a} X E_{a,a}^*, \tag{2.167}$$

and an example of a Stinespring representation of Δ is

$$\Delta(X) = \mathrm{Tr}_{\mathcal{Z}}\left(AXA^*\right) \tag{2.168}$$

for $\mathcal{Z} = \mathbb{C}^{\Sigma}$ and

$$A = \sum_{a \in \Sigma} (e_a \otimes e_a) e_a^*. \tag{2.169}$$

A digression on classical registers

Classical probabilistic states of registers may be associated with diagonal density operators, as discussed in Section 2.1.2. The term *classical register* was mentioned in that discussion but not fully explained, as channels had not yet been introduced at that point. It is appropriate to make this notion more precise, now that channels (and the completely dephasing channel in particular) have been introduced.

From a mathematical point of view, classical registers are not defined in a manner that is distinct from ordinary (quantum) registers. Rather, the term *classical register* will be used to refer to any register that, by the nature of the processes under consideration, would be unaffected by an application of the completely dephasing channel Δ at any moment during its existence. Every state of a classical register at must be a diagonal density operator, corresponding to a probabilistic state, as these are the density operators that are invariant under the action of the channel Δ. Moreover, the correlations that may exist between a classical register and one or more other registers is limited. For example, for a classical register X and an arbitrary register Y, the only states of the compound register (X, Y) that are consistent with the term *classical register* being applied to X are those taking the form

$$\sum_{a \in \Sigma} p(a) E_{a,a} \otimes \rho_a, \tag{2.170}$$

for Σ being the classical state set of X, $\{\rho_a : a \in \Sigma\} \subset \mathrm{D}(\mathcal{X})$ being an arbitrary collection of states of Y, and $p \in \mathcal{P}(\Sigma)$ being a probability vector. States of this form are commonly called *classical-quantum* states. It is both natural and convenient in some situations to associate the state (2.170) with the ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined as $\eta(a) = p(a)\rho_a$ for each $a \in \Sigma$.

2.2.4 Extremal channels

For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, the set of quantum channels $\mathrm{C}(\mathcal{X},\mathcal{Y})$ is compact and convex (by Proposition 2.28). A characterization of the extreme points of this set is given by Theorem 2.31 below. The following lemma will be used in the proof of this theorem.

Lemma 2.30. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $A \in \mathrm{L}(\mathcal{Y},\mathcal{X})$ be an operator. It holds that*

$$\bigl\{P \in \mathrm{Pos}(\mathcal{X}) \,:\, \mathrm{im}(P) \subseteq \mathrm{im}(A)\bigr\} = \bigl\{AQA^* \,:\, Q \in \mathrm{Pos}(\mathcal{Y})\bigr\}. \tag{2.171}$$

Proof. For every $Q \in \mathrm{Pos}(\mathcal{Y})$, it holds that AQA^* is positive semidefinite and satisfies $\mathrm{im}(AQA^*) \subseteq \mathrm{im}(A)$. The set on the right-hand side of (2.171) is therefore contained in the set on the left-hand side.

For the reverse containment, if $P \in \mathrm{Pos}(\mathcal{X})$ satisfies $\mathrm{im}(P) \subseteq \mathrm{im}(A)$, then by setting

$$Q = A^{+}P(A^{+})^*, \tag{2.172}$$

for A^{+} denoting the Moore–Penrose pseudo-inverse of A, one obtains

$$AQA^* = (AA^{+})P(AA^{+})^* = \Pi_{\mathrm{im}(A)}P\Pi_{\mathrm{im}(A)} = P, \tag{2.173}$$

which completes the proof. □

Theorem 2.31 (Choi). *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel, and let $\{A_a \,:\, a \in \Sigma\} \subset \mathrm{L}(\mathcal{X},\mathcal{Y})$ be a linearly independent set of operators satisfying*

$$\Phi(X) = \sum_{a\in\Sigma} A_a X A_a^* \tag{2.174}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The channel Φ is an extreme point of the set $\mathrm{C}(\mathcal{X},\mathcal{Y})$ if and only if the collection

$$\bigl\{A_b^* A_a \,:\, (a,b) \in \Sigma\times\Sigma\bigr\} \subset \mathrm{L}(\mathcal{X}) \tag{2.175}$$

of operators is a linearly independent.

Proof. Let $\mathcal{Z} = \mathbb{C}^{\Sigma}$, define an operator $M \in \mathrm{L}(\mathcal{Z},\mathcal{Y}\otimes\mathcal{X})$ as

$$M = \sum_{a\in\Sigma} \mathrm{vec}(A_a)e_a^*, \tag{2.176}$$

and observe that

$$MM^* = \sum_{a\in\Sigma} \mathrm{vec}(A_a)\,\mathrm{vec}(A_a)^* = J(\Phi). \tag{2.177}$$

As $\{A_a : a \in \Sigma\}$ is a linearly independent collection of operators, it must hold that $\ker(M) = \{0\}$.

Assume first that Φ is not an extreme point of $\mathrm{C}(\mathcal{X}, \mathcal{Y})$. It follows that there exist channels $\Psi_0, \Psi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, with $\Psi_0 \neq \Psi_1$, along with a scalar $\lambda \in (0, 1)$, such that

$$\Phi = \lambda \Psi_0 + (1 - \lambda)\Psi_1. \tag{2.178}$$

Let $P = J(\Phi)$, $Q_0 = J(\Psi_0)$, and $Q_1 = J(\Psi_1)$, so that

$$P = \lambda Q_0 + (1 - \lambda)Q_1. \tag{2.179}$$

As Φ, Ψ_0, and Ψ_1 are channels, the operators $P, Q_0, Q_1 \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$ are positive semidefinite and satisfy

$$\mathrm{Tr}_{\mathcal{Y}}(P) = \mathrm{Tr}_{\mathcal{Y}}(Q_0) = \mathrm{Tr}_{\mathcal{Y}}(Q_1) = \mathbb{1}_{\mathcal{X}}, \tag{2.180}$$

by Corollary 2.27.

Because λ is positive and the operators Q_0 and Q_1 are positive semidefinite, the equation (2.179) implies

$$\mathrm{im}(Q_0) \subseteq \mathrm{im}(P) = \mathrm{im}(M). \tag{2.181}$$

It follows by Lemma 2.30 that there exists a positive semidefinite operator $R_0 \in \mathrm{Pos}(\mathcal{Z})$ for which $Q_0 = MR_0M^*$. By similar reasoning, there exists a positive semidefinite operator $R_1 \in \mathrm{Pos}(\mathcal{Z})$ for which $Q_1 = MR_1M^*$.

Letting $H = R_0 - R_1$, one finds that

$$0 = \mathrm{Tr}_{\mathcal{Y}}(Q_0) - \mathrm{Tr}_{\mathcal{Y}}(Q_1) = \mathrm{Tr}_{\mathcal{Y}}\big(MHM^*\big) = \sum_{a,b \in \Sigma} H(a,b)\big(A_b^* A_a\big)^{\mathsf{T}}, \tag{2.182}$$

and therefore

$$\sum_{a,b \in \Sigma} H(a,b) A_b^* A_a = 0. \tag{2.183}$$

Because $\Psi_0 \neq \Psi_1$, it holds that $Q_0 \neq Q_1$, so $R_0 \neq R_1$, and therefore $H \neq 0$. It has therefore been proved that $\big\{A_b^* A_a : (a,b) \in \Sigma \times \Sigma\big\}$ is a linearly dependent collection of operators.

Now assume the set (2.175) is linearly dependent:

$$\sum_{a,b \in \Sigma} Z(a,b) A_b^* A_a = 0 \tag{2.184}$$

for some choice of a nonzero operator $Z \in \mathrm{L}(\mathcal{Z})$. It follows that

$$\sum_{a,b \in \Sigma} H(a,b) A_b^* A_a = 0 \tag{2.185}$$

for both of the Hermitian operators

$$H = \frac{Z + Z^*}{2} \quad \text{and} \quad H = \frac{Z - Z^*}{2i}. \tag{2.186}$$

At least one of these operators must be nonzero, which implies that (2.185) must hold for some choice of a nonzero Hermitian operator H. Let such a choice of H be fixed, and define $K = H/\|H\|$.

Let $\Psi_0, \Psi_1 \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be the mappings defined by the equations

$$J(\Psi_0) = M(\mathbb{1} + K)M^* \quad \text{and} \quad J(\Psi_1) = M(\mathbb{1} - K)M^*. \tag{2.187}$$

Because K is Hermitian and satisfies $\|K\| \leq 1$, one has that the operators $\mathbb{1} + K$ and $\mathbb{1} - K$ are both positive semidefinite. The operators $M(\mathbb{1} + K)M^*$ and $M(\mathbb{1} - K)M^*$ are therefore positive semidefinite as well, implying that Ψ_0 and Ψ_1 are completely positive, by Theorem 2.22. It holds that

$$\begin{aligned} \mathrm{Tr}_{\mathcal{Y}}\left(MHM^*\right) &= \sum_{a,b \in \Sigma} H(a,b) \left(A_b^* A_a\right)^{\mathsf{T}} \\ &= \left(\sum_{a,b \in \Sigma} H(a,b) A_b^* A_a \right)^{\mathsf{T}} = 0 \end{aligned} \tag{2.188}$$

and therefore the following two equations hold:

$$\begin{aligned} \mathrm{Tr}_{\mathcal{Y}}\left(J(\Psi_0)\right) &= \mathrm{Tr}_{\mathcal{Y}}\left(MM^*\right) + \mathrm{Tr}_{\mathcal{Y}}\left(MHM^*\right) = \mathrm{Tr}_{\mathcal{Y}}(J(\Phi)) = \mathbb{1}_{\mathcal{X}}, \\ \mathrm{Tr}_{\mathcal{Y}}\left(J(\Psi_1)\right) &= \mathrm{Tr}_{\mathcal{Y}}\left(MM^*\right) - \mathrm{Tr}_{\mathcal{Y}}\left(MHM^*\right) = \mathrm{Tr}_{\mathcal{Y}}(J(\Phi)) = \mathbb{1}_{\mathcal{X}}. \end{aligned} \tag{2.189}$$

Thus, Ψ_0 and Ψ_1 are trace-preserving by Theorem 2.26, and are therefore channels.

Finally, given that $H \neq 0$ and $\ker(M) = \{0\}$, it holds that $J(\Psi_0) \neq J(\Psi_1)$, so that $\Psi_0 \neq \Psi_1$. As

$$\frac{1}{2} J(\Psi_0) + \frac{1}{2} J(\Psi_1) = MM^* = J(\Phi), \tag{2.190}$$

one has that

$$\Phi = \frac{1}{2}\Psi_0 + \frac{1}{2}\Psi_1, \tag{2.191}$$

which demonstrates that Φ is not an extreme point of $\mathrm{C}(\mathcal{X}, \mathcal{Y})$. □

Example 2.32. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, assume $\dim(\mathcal{X}) \leq \dim(\mathcal{Y})$, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$ be an isometry, and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be the isometric channel defined by

$$\Phi(X) = AXA^* \tag{2.192}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The set $\{A^*A\}$ contains a single nonzero operator, and is therefore linearly independent. By Theorem 2.31, Φ is an extreme point of the set $\mathrm{C}(\mathcal{X}, \mathcal{Y})$.

Example 2.33. Let $\Sigma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{X} = \mathbb{C}^{\Sigma}$ and $\mathcal{Y} = \mathbb{C}^{\Sigma \times \Sigma}$. Also define operators $A_0, A_1 \in \mathrm{L}(\mathcal{X}, \mathcal{Y})$ as

$$\begin{aligned} A_0 &= \frac{1}{\sqrt{6}}\bigl(2E_{00,0} + E_{01,1} + E_{10,1}\bigr), \\ A_1 &= \frac{1}{\sqrt{6}}\bigl(2E_{11,1} + E_{01,0} + E_{10,0}\bigr), \end{aligned} \tag{2.193}$$

(with elements of the form $(a, b) \in \Sigma \times \Sigma$ being written as ab for the sake of clarity). Expressed as matrices (with respect to the natural orderings of Σ and $\Sigma \times \Sigma$), these operators are as follows:

$$A_0 = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad A_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 2 \end{pmatrix}. \tag{2.194}$$

Now, define a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ as

$$\Phi(X) = A_0 X A_0^* + A_1 X A_1^* \tag{2.195}$$

for every $X \in \mathrm{L}(\mathcal{X})$. It holds that

$$\begin{aligned} A_0^*A_0 &= \frac{1}{3}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A_0^*A_1 = \frac{1}{3}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \\ A_1^*A_0 &= \frac{1}{3}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad A_1^*A_1 = \frac{1}{3}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}. \end{aligned} \tag{2.196}$$

The set $\{A_0^*A_0, A_0^*A_1, A_1^*A_0, A_1^*A_1\}$ is linearly independent, and therefore Theorem 2.31 implies that Φ is an extreme point of $\mathrm{C}(\mathcal{X}, \mathcal{Y})$.

2.3 Measurements

Measurements provide the mechanism through which classical information may be extracted from quantum states. This section defines measurements, and various notions connected with measurements, and provides a basic mathematical development of this concept.

2.3.1 Two equivalent definitions of measurements

When a hypothetical observer measures a register, the observer does not see a description of that register's quantum state as a density operator, but instead obtains a classical measurement outcome. In general, this outcome is understood to be sampled randomly according to some probability distribution, which is determined by the measurement together with the quantum state of the register immediately before the measurement was performed. In this way, measurements allow one to associate a meaning to the density operator description of quantum states, at least insofar as the density operators determine the probabilities with which different classical outcomes occur for each possible measurement.

Measurements can be defined in mathematical terms in two different, but equivalent, ways. Both ways will be described in this section, and their equivalence will be explained.

Measurements defined by measurement operators

The following definition represents the first formulation of measurements to be described in this book. The precise mathematical meaning of the term *measurement* used throughout this book coincides with this definition.

Definition 2.34. A *measurement* is a function of the form

$$\mu : \Sigma \to \mathrm{Pos}(\mathcal{X}), \tag{2.197}$$

for some choice of an alphabet Σ and a complex Euclidean space $\mathcal{X}$, satisfying the constraint

$$\sum_{a \in \Sigma} \mu(a) = \mathbb{1}_{\mathcal{X}}. \tag{2.198}$$

The set Σ is the set of *measurement outcomes* of this measurement, and each operator $\mu(a)$ is the *measurement operator* associated with the corresponding outcome $a \in \Sigma$.

When a measurement μ is performed on a given register X, it must be assumed that μ takes the form (2.197), for an arbitrary choice of an alphabet Σ and for $\mathcal{X}$ being the complex Euclidean space associated with X. Two things happen when such a measurement is performed, assuming the state of X immediately prior to the measurement is $\rho \in \mathrm{D}(\mathcal{X})$:

1. An element of Σ is selected at random. The probability distribution that describes this random selection is represented by the probability vector $p \in \mathcal{P}(\Sigma)$ defined as
$$p(a) = \langle \mu(a), \rho \rangle \tag{2.199}$$
for each $a \in \Sigma$.
2. The register X ceases to exist, in the sense that it no longer has a defined state and cannot be considered in further calculations.

It is evident from the first item that the probabilities associated with the outcomes of a given measurement depend linearly on the state that is measured. It is also evident that the probability vector $p \in \mathcal{P}(\Sigma)$ defined by (2.199) is indeed a probability vector: as ρ and $\mu(a)$ are both positive semidefinite, their inner product $p(a) = \langle \mu(a), \rho \rangle$ is nonnegative, and summing these values gives
$$\sum_{a \in \Sigma} p(a) = \sum_{a \in \Sigma} \langle \mu(a), \rho \rangle = \langle \mathbb{1}_{\mathcal{X}}, \rho \rangle = \mathrm{Tr}(\rho) = 1. \tag{2.200}$$

The assumption that registers cease to exist after being measured is not universal within quantum information theory—an alternative definition, in which the states of registers after they are measured is specified, does not make this requirement. Measurements of this alternative type, which are called *nondestructive measurements* in this book, are discussed in greater detail in Section 2.3.2. It is the case that nondestructive measurements can be described as compositions of ordinary (destructive) measurements and channels, and need not be considered as fundamental objects within the theory for this reason.

It is sometimes convenient to specify a measurement by describing its measurement operators as a collection indexed by its set of measurement outcomes. In particular, when one refers to a measurement as a collection
$$\{P_a : a \in \Sigma\} \subset \mathrm{Pos}(\mathcal{X}), \tag{2.201}$$
it is to be understood that the measurement is given by $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, where $\mu(a) = P_a$ for each $a \in \Sigma$.

Measurements as channels

The second formulation of measurements, which is equivalent to the first, essentially describes measurements as channels whose outputs are stored in classical registers. The following definition of *quantum-to-classical channels* makes this notion precise.

Definition 2.35. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$. It is said that Φ is a *quantum-to-classical channel* if and only if

$$\Phi = \Delta\Phi, \tag{2.202}$$

for $\Delta \in \mathrm{C}(\mathcal{Y})$ denoting the completely dephasing channel, defined with respect to the space $\mathcal{Y}$.

An equivalent condition for a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ to be a quantum-to-classical channel is that $\Phi(\rho)$ is a diagonal density operator for every $\rho \in \mathrm{D}(\mathcal{X})$. The following simple proposition establishes that this is so.

Proposition 2.36. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that Φ is a quantum-to-classical channel if and only if $\Phi(\rho)$ is diagonal for every $\rho \in \mathrm{D}(\mathcal{X})$.*

Proof. If Φ is a quantum-to-classical channel, then

$$\Phi(\rho) = \Delta(\Phi(\rho)), \tag{2.203}$$

and therefore $\Phi(\rho)$ is diagonal, for every density operator $\rho \in \mathrm{D}(\mathcal{X})$.

Conversely, if $\Phi(\rho)$ is diagonal, then $\Phi(\rho) = \Delta(\Phi(\rho))$, and therefore

$$(\Phi - \Delta\Phi)(\rho) = 0, \tag{2.204}$$

for every $\rho \in \mathrm{D}(\mathcal{X})$. As the density operators $\mathrm{D}(\mathcal{X})$ span all of $\mathrm{L}(\mathcal{X})$, it follows that $\Phi = \Delta\Phi$, and therefore Φ is a quantum-to-classical channel. □

The equivalence between measurements and quantum-to-classical channels is revealed by the following theorem. In essence, quantum-to-classical channels of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ represent precisely those channels that can be realized as a measurement of a register X, according to some measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, followed by the measurement outcome being stored in a register Y having classical state set Σ.

Theorem 2.37. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\mathcal{Y} = \mathbb{C}^{\Sigma}$. The following two complementary facts hold:*

1. *For every quantum-to-classical channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, there exists a uniquely determined measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ for which the equation*

$$\Phi(X) = \sum_{a \in \Sigma} \langle \mu(a), X \rangle \, E_{a,a} \tag{2.205}$$

   holds for all $X \in \mathrm{L}(\mathcal{X})$.

2. *For every measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, the mapping $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ defined by (2.205) for all $X \in \mathrm{L}(\mathcal{X})$ is a quantum-to-classical channel.*

Proof. Assume first that $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is a quantum-to-classical channel. It therefore holds that

$$\Phi(X) = \Delta(\Phi(X)) = \sum_{a \in \Sigma} \langle E_{a,a}, \Phi(X) \rangle \, E_{a,a} = \sum_{a \in \Sigma} \langle \Phi^*(E_{a,a}), X \rangle \, E_{a,a} \tag{2.206}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Define a function $\mu : \Sigma \to \mathrm{L}(\mathcal{X})$ as

$$\mu(a) = \Phi^*(E_{a,a}) \tag{2.207}$$

for each $a \in \Sigma$. As Φ is completely positive, so too is Φ^* (as explained in Remark 2.18), and therefore $\mu(a) \in \mathrm{Pos}(\mathcal{X})$ for each $a \in \Sigma$. Moreover, as Φ is trace-preserving, it holds (by Theorem 2.26) that Φ^* is unital, and therefore

$$\sum_{a \in \Sigma} \mu(a) = \sum_{a \in \Sigma} \Phi^*(E_{a,a}) = \Phi^*(\mathbb{1}_{\mathcal{Y}}) = \mathbb{1}_{\mathcal{X}}. \tag{2.208}$$

It follows that μ is a measurement for which (2.205) holds for all $X \in \mathrm{L}(\mathcal{X})$.

Toward proving the uniqueness of the measurement μ satisfying (2.205) for all $X \in \mathrm{L}(\mathcal{X})$, let $\nu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an arbitrary measurement for which the equation

$$\Phi(X) = \sum_{a \in \Sigma} \langle \nu(a), X \rangle \, E_{a,a} \tag{2.209}$$

holds for all $X \in \mathrm{L}(\mathcal{X})$. One then has that

$$\sum_{a \in \Sigma} \langle \mu(a) - \nu(a), X \rangle \, E_{a,a} = 0 \tag{2.210}$$

for all $X \in \mathrm{L}(\mathcal{X})$, which implies that $\nu(a) = \mu(a)$ for every $a \in \Sigma$, and completes the proof of the first fact.

Now assume that $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ is a measurement, and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be defined by (2.205). The Choi representation of this mapping is

$$J(\Phi) = \sum_{a \in \Sigma} E_{a,a} \otimes \overline{\mu(a)}. \tag{2.211}$$

This is a positive semidefinite operator, and it holds that

$$\mathrm{Tr}_{\mathcal{Y}}(J(\Phi)) = \sum_{a \in \Sigma} \overline{\mu(a)} = \overline{\mathbb{1}_{\mathcal{X}}} = \mathbb{1}_{\mathcal{X}}. \tag{2.212}$$

By Corollary 2.27, it holds that Φ is a channel. It is evident from inspection that $\Phi(\rho)$ is diagonal for every $\rho \in \mathrm{D}(\mathcal{X})$, and therefore Φ is a quantum-to-classical channel by Proposition 2.36, which completes the proof of the second statement. □

As the following proposition establishes, the set of quantum-to-classical channels of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is both compact and convex.

Proposition 2.38. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces. The set of quantum-to-classical channels of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is compact and convex.*

Proof. It will first be observed that the set of all quantum-to-classical channels of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is given by

$$\big\{\Delta\Psi \,:\, \Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})\big\}, \tag{2.213}$$

for $\Delta \in \mathrm{C}(\mathcal{Y})$ being the completely dephasing channel defined with respect to the space $\mathcal{Y}$. Indeed, for every channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, it holds that $\Delta\Psi$ is a quantum-to-classical channel by virtue of the fact that the channel Δ is idempotent (i.e., $\Delta\Delta = \Delta$). On the other hand, every quantum-to-classical channel Φ satisfies $\Phi = \Delta\Phi$ by definition, and is therefore represented in the set (2.213) by taking $\Psi = \Phi$.

By Proposition 2.28, it holds that the set $\mathrm{C}(\mathcal{X}, \mathcal{Y})$ is compact and convex. As the mapping $\Psi \mapsto \Delta\Psi$ defined on $\mathrm{C}(\mathcal{X}, \mathcal{Y})$ is linear (and therefore continuous, as the dimension of $\mathrm{T}(\mathcal{X}, \mathcal{Y})$ is finite), it must map $\mathrm{C}(\mathcal{X}, \mathcal{Y})$ to a compact and convex set. The image of $\mathrm{C}(\mathcal{X}, \mathcal{Y})$ under this mapping is precisely the set (2.213), which coincides with the set of quantum-to-classical channels of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, so the proof is complete. □

2.3.2 Basic notions concerning measurements

The subsections that follow introduce various notions and facts connected with measurements.

Product measurements

Suppose $\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ is a compound register. One may then consider a collection of measurements

$$\begin{aligned} \mu_1 &: \Sigma_1 \to \mathrm{Pos}(\mathcal{Y}_1) \\ &\vdots \\ \mu_n &: \Sigma_n \to \mathrm{Pos}(\mathcal{Y}_n) \end{aligned} \tag{2.214}$$

to be performed independently on the registers $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$. Such a process may be viewed as a single measurement

$$\mu : \Sigma_1 \times \cdots \times \Sigma_n \to \mathrm{Pos}(\mathcal{X}) \tag{2.215}$$

on X that is defined as

$$\mu(a_1, \ldots, a_n) = \mu_1(a_1) \otimes \cdots \otimes \mu_n(a_n) \tag{2.216}$$

for each tuple $(a_1, \ldots, a_n) \in \Sigma_1 \times \cdots \times \Sigma_n$. A measurement μ of this sort is said to be a *product measurement* on X.

It may be verified that when a product measurement is performed on a product state, the measurement outcomes resulting from the individual measurements are independently distributed.

Partial measurements

Suppose $\mathsf{X} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n)$ is a compound register, and a measurement

$$\mu : \Sigma \to \mathrm{Pos}(\mathcal{Y}_k) \tag{2.217}$$

is performed only on the register Y_k, for a single choice of $k \in \{1, \ldots, n\}$. Such a measurement must not only produce a measurement outcome $a \in \Sigma$, but must also determine the resulting state of the register

$$(\mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n), \tag{2.218}$$

conditioned on the measurement outcome that was obtained. For a given state $\rho \in \mathrm{D}(\mathcal{X})$ of the register X, the probability for each measurement outcome to appear, along with the corresponding post-measurement state of the register (2.218), may be calculated by considering the quantum-to-classical channel that corresponds to the measurement μ.

Let this quantum-to-channel be denoted by $\Phi \in \mathrm{C}(\mathcal{Y}_k, \mathcal{Z})$, for $\mathcal{Z} = \mathbb{C}^{\Sigma}$, so that

$$\Phi(Y) = \sum_{a \in \Sigma} \langle \mu(a), Y \rangle \, E_{a,a} \tag{2.219}$$

for every $Y \in \mathrm{L}(\mathcal{Y}_k)$. Consider the state of the compound register

$$(\mathsf{Z}, \mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n) \tag{2.220}$$

obtained by applying the channel Φ to Y_k, followed by the application of a channel that performs the permutation of registers

$$(\mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Z}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n) \to (\mathsf{Z}, \mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n) \tag{2.221}$$

without changing the contents of these individual registers. The state of the register (2.220) that results may be written explicitly as

$$\sum_{a \in \Sigma} E_{a,a} \otimes \mathrm{Tr}_{\mathcal{Y}_k}\big((\mathbb{1}_{\mathcal{Y}_1} \otimes \cdots \otimes \mathbb{1}_{\mathcal{Y}_{k-1}} \otimes \mu(a) \otimes \mathbb{1}_{\mathcal{Y}_{k+1}} \otimes \cdots \otimes \mathbb{1}_{\mathcal{Y}_n})\rho\big). \tag{2.222}$$

The state (2.222) is a classical-quantum state, and is naturally associated with the ensemble

$$\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_{k-1} \otimes \mathcal{Y}_{k+1} \otimes \cdots \otimes \mathcal{Y}_n) \tag{2.223}$$

defined as

$$\eta(a) = \mathrm{Tr}_{\mathcal{Y}_k}\big((\mathbb{1}_{\mathcal{Y}_1} \otimes \cdots \otimes \mathbb{1}_{\mathcal{Y}_{k-1}} \otimes \mu(a) \otimes \mathbb{1}_{\mathcal{Y}_{k+1}} \otimes \cdots \otimes \mathbb{1}_{\mathcal{Y}_n})\rho\big) \tag{2.224}$$

for each measurement outcome $a \in \Sigma$. This ensemble describes both the distribution of measurement outcomes of the measurement μ, together with the resulting states of the remaining registers. Equivalently, each measurement outcome $a \in \Sigma$ appears with probability

$$\mathrm{Tr}(\eta(a)) = \big\langle \mu(a), \rho[\mathsf{Y}_k] \big\rangle, \tag{2.225}$$

and conditioned on each outcome $a \in \Sigma$ that appears with a nonzero probability, the resulting state of $(\mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1}, \mathsf{Y}_{k+1}, \ldots, \mathsf{Y}_n)$ becomes

$$\frac{\mathrm{Tr}_{\mathcal{Y}_k}\big((\mathbb{1}_{\mathcal{Y}_1} \otimes \cdots \otimes \mathbb{1}_{\mathcal{Y}_{k-1}} \otimes \mu(a) \otimes \mathbb{1}_{\mathcal{Y}_{k+1}} \otimes \cdots \otimes \mathbb{1}_{\mathcal{Y}_n})\rho\big)}{\big\langle \mu(a), \rho[\mathsf{Y}_k] \big\rangle}. \tag{2.226}$$

Example 2.39. Let Σ be an alphabet, and let Y and Z be registers whose classical state sets are given by Σ, so that $\mathcal{Y} = \mathbb{C}^{\Sigma}$ and $\mathcal{Z} = \mathbb{C}^{\Sigma}$. Define a state $\tau \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$ as

$$\tau = \frac{1}{|\Sigma|} \sum_{b,c \in \Sigma} E_{b,c} \otimes E_{b,c}\,, \tag{2.227}$$

and consider an arbitrary measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ on $\mathcal{Y}$. If this measurement is performed on Y when the pair (Y, Z) is in the state τ, then each outcome $a \in \Gamma$ appears with probability

$$p(a) = \big\langle \mu(a), \rho[\mathsf{Y}] \big\rangle = \frac{\mathrm{Tr}(\mu(a))}{|\Sigma|}. \tag{2.228}$$

Conditioned on the event that the measurement outcome a appears, the state of Z becomes

$$\begin{aligned} &\frac{1}{p(a)} \mathrm{Tr}_{\mathcal{Y}}\big((\mu(a) \otimes \mathbb{1}_{\mathcal{Z}})\tau\big) \\ &\quad = \frac{|\Sigma|}{\mathrm{Tr}(\mu(a))} \frac{1}{|\Sigma|} \sum_{b,c \in \Sigma} \langle \mu(a), E_{b,c} \rangle E_{b,c} = \frac{\overline{\mu(a)}}{\mathrm{Tr}(\mu(a))}. \end{aligned} \tag{2.229}$$

Projective measurements and Naimark's theorem

A measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ is said to be a *projective measurement* if and only if each of its measurement operators is a projection: $\mu(a) \in \mathrm{Proj}(\mathcal{X})$ for every $a \in \Sigma$.

The following simple proposition demonstrates that the measurement operators of a projective measurement must be pairwise orthogonal, and must therefore project onto orthogonal subspaces. For a given projective measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, there can therefore be no more than $\dim(\mathcal{X})$ distinct values of $a \in \Sigma$ for which $\mu(a)$ is nonzero.

Proposition 2.40. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a projective measurement. The set $\{\mu(a) : a \in \Sigma\}$ is an orthogonal set.*

Proof. As μ is a measurement, it holds that

$$\sum_{a \in \Sigma} \mu(a) = \mathbb{1}_{\mathcal{X}}, \tag{2.230}$$

and therefore this sum must square to itself:

$$\sum_{a,b\in\Sigma} \mu(a)\mu(b) = \left(\sum_{a\in\Sigma} \mu(a)\right)^2 = \sum_{a\in\Sigma} \mu(a). \tag{2.231}$$

Because each operator $\mu(a)$ is a projection operator, it follows that

$$\sum_{a,b\in\Sigma} \mu(a)\mu(b) = \sum_{a\in\Sigma} \mu(a) + \sum_{\substack{a,b\in\Sigma \\ a\neq b}} \mu(a)\mu(b), \tag{2.232}$$

and therefore

$$\sum_{\substack{a,b\in\Sigma \\ a\neq b}} \mu(a)\mu(b) = 0. \tag{2.233}$$

Taking the trace of both sides of this equation yields

$$\sum_{\substack{a,b\in\Sigma \\ a\neq b}} \langle \mu(a), \mu(b) \rangle = 0. \tag{2.234}$$

The inner product of any two positive semidefinite operators is nonnegative, and therefore $\langle \mu(a), \mu(b) \rangle = 0$ for all $a, b \in \Sigma$ with $a \neq b$, which completes the proof. □

For any orthonormal basis $\{x_a : a \in \Sigma\}$ of a complex Euclidean space $\mathcal{X} = \mathbb{C}^\Sigma$, the measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined as

$$\mu(a) = x_a x_a^* \tag{2.235}$$

for each $a \in \Sigma$ is an example of a projective measurement. Such a measurement is, more specifically, known as a *complete projective measurement*. This is the measurement that is commonly referred to as the *measurement with respect to the basis* $\{x_a : a \in \Sigma\}$.

Example 2.41. Let Σ be an alphabet and let $\mathcal{X} = \mathbb{C}^\Sigma$. The *measurement with respect to the standard basis* of $\mathcal{X}$ is the measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined as

$$\mu(a) = E_{a,a} \tag{2.236}$$

for each $a \in \Sigma$. For a given state $\rho \in \mathrm{D}(\mathcal{X})$, the probability associated with each measurement outcome $a \in \Sigma$, were this state to be measured according to μ, is equal to the corresponding diagonal entry $\rho(a, a)$. One may also observe that the quantum-to-classical channel associated with this measurement is the completely dephasing channel $\Delta \in \mathrm{C}(\mathcal{X})$.

The following theorem, known as *Naimark's theorem*, establishes a link between arbitrary measurements and projective measurements. It implies that any measurement can be viewed as a projective measurement on a compound register that includes the original register as a subregister.

Theorem 2.42 (Naimark's theorem)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a measurement, and let $\mathcal{Y} = \mathbb{C}^{\Sigma}$. There exists an isometry $A \in \mathrm{U}\big(\mathcal{X}, \mathcal{X} \otimes \mathcal{Y}\big)$ such that*

$$\mu(a) = A^*(\mathbb{1}_{\mathcal{X}} \otimes E_{a,a})A \tag{2.237}$$

for every $a \in \Sigma$.

Proof. Define an operator $A \in \mathrm{L}\big(\mathcal{X}, \mathcal{X} \otimes \mathcal{Y}\big)$ as

$$A = \sum_{a \in \Sigma} \sqrt{\mu(a)} \otimes e_a. \tag{2.238}$$

It holds that

$$A^*A = \sum_{a \in \Sigma} \mu(a) = \mathbb{1}_{\mathcal{X}}, \tag{2.239}$$

and therefore A is an isometry. The required equation (2.237) holds for each $a \in \Sigma$, so the proof is complete. □

Corollary 2.43. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a measurement. Also let $\mathcal{Y} = \mathbb{C}^{\Sigma}$ and let $u \in \mathcal{Y}$ be a unit vector. There exists a projective measurement $\nu : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ such that*

$$\langle \nu(a), X \otimes uu^* \rangle = \langle \mu(a), X \rangle \tag{2.240}$$

for every $X \in \mathrm{L}(\mathcal{X})$.

Proof. Let $A \in \mathrm{U}(\mathcal{X}, \mathcal{X} \otimes \mathcal{Y})$ be the isometry whose existence is implied by Theorem 2.42. Choose $U \in \mathrm{U}\big(\mathcal{X} \otimes \mathcal{Y}\big)$ to be any unitary operator for which the equation

$$U(\mathbb{1}_{\mathcal{X}} \otimes u) = A \tag{2.241}$$

is satisfied, and define $\nu : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ as

$$\nu(a) = U^*(\mathbb{1}_{\mathcal{X}} \otimes E_{a,a})U \tag{2.242}$$

for each $a \in \Sigma$. It holds that ν is a projective measurement, and moreover

$$\begin{aligned} \langle \nu(a), X \otimes uu^* \rangle &= \langle (\mathbb{1}_{\mathcal{X}} \otimes u^*)U^*(\mathbb{1}_{\mathcal{X}} \otimes E_{a,a})U(\mathbb{1}_{\mathcal{X}} \otimes u), X \rangle \\ &= \langle A^*(\mathbb{1}_{\mathcal{X}} \otimes E_{a,a})A, X \rangle = \langle \mu(a), X \rangle \end{aligned} \tag{2.243}$$

for each $a \in \Sigma$, as required. □

Information-complete measurements

States of registers are uniquely determined by the measurement statistics they generate. More precisely, the knowledge of the probability associated with every outcome of every measurement that could be performed on a given register is sufficient to obtain a description of that register's state. In fact, something stronger may be said, which is that there exist choices of measurements that uniquely determine every possible state of a register by the measurement statistics that they alone generate. Such measurements, which are known as *information-complete measurements*, are characterized by the property that their measurement operators span the space of operators from which they are drawn.

In more explicit terms, a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ on a complex Euclidean space $\mathcal{X}$ is said to be an *information-complete* measurement if it holds that

$$\mathrm{span}\bigl\{\mu(a) \,:\, a \in \Sigma\bigr\} = \mathrm{L}(\mathcal{X}). \tag{2.244}$$

For any such measurement, and for any choice of $\rho \in \mathrm{D}(\mathcal{X})$, it holds that the probability vector $p \in \mathcal{P}(\Sigma)$ defined by $p(a) = \langle \mu(a), \rho \rangle$ uniquely determines ρ. This fact is evident from the following proposition.

Proposition 2.44. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\{A_a \,:\, a \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$ be a collection of operators that spans $\mathrm{L}(\mathcal{X})$. The mapping $\phi : \mathrm{L}(\mathcal{X}) \to \mathbb{C}^{\Sigma}$ defined by*

$$(\phi(X))(a) = \langle A_a, X \rangle, \tag{2.245}$$

for each $X \in \mathrm{L}(\mathcal{X})$ and $a \in \Sigma$, is an injective mapping.

Proof. Let $X, Y \in \mathrm{L}(\mathcal{X})$ satisfy $\phi(X) = \phi(Y)$, so that

$$\langle A_a, X - Y \rangle = 0 \tag{2.246}$$

for every $a \in \Sigma$. As $\{A_a \,:\, a \in \Sigma\}$ spans $\mathrm{L}(\mathcal{X})$, it follows that

$$\langle Z, X - Y \rangle = 0 \tag{2.247}$$

for every $Z \in \mathrm{L}(\mathcal{X})$, and consequently $X - Y = 0$, which completes the proof. □

The following example provides one way of constructing information-complete measurements, for any choice of a complex Euclidean space.

Example 2.45. Let Σ be an alphabet, let $\mathcal{X} = \mathbb{C}^{\Sigma}$, and let

$$\{\rho_{a,b} \,:\, (a,b) \in \Sigma \times \Sigma\} \subset \mathrm{D}(\mathcal{X}) \tag{2.248}$$

be a collection of density operators that spans all of $\mathrm{L}(\mathcal{X})$. One such set was constructed in Example 2.6. Also define

$$Q = \sum_{(a,b)\in\Sigma\times\Sigma} \rho_{a,b} \tag{2.249}$$

and observe that Q is necessarily positive definite; if this were not so, there would exist a nonzero vector $u \in \mathcal{X}$ satisfying $\langle \rho_{a,b}, uu^* \rangle = 0$ for each pair $(a,b) \in \Sigma \times \Sigma$, in contradiction with Proposition 2.44. It may be verified that the function $\mu : \Sigma \times \Sigma \to \mathrm{Pos}(\mathcal{X})$, defined by

$$\mu(a,b) = Q^{-\frac{1}{2}} \rho_{a,b} Q^{-\frac{1}{2}} \tag{2.250}$$

for each $(a,b) \in \Sigma \times \Sigma$, is an information-complete measurement.

Nondestructive measurements and quantum instruments

It is convenient in some situations to consider an alternative definition of measurements that does not dictate that registers are destroyed upon being measured. Instead, a measured register is left in some particular state that depends both on its initial state and on the measurement outcome obtained. More generally, one may consider that the measured register is transformed into another register as a result of the measurement process.

One specific definition, which is frequently taken as the definition of a measurement by other authors, describes such a process as a function of the form $\nu : \Sigma \to \mathrm{L}(\mathcal{X})$, often expressed as a collection

$$\{M_a \,:\, a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}), \tag{2.251}$$

where Σ is the alphabet of measurement outcomes and $\mathcal{X}$ is the complex Euclidean space corresponding to the register being measured, such that the constraint

$$\sum_{a\in\Sigma} M_a^* M_a = \mathbb{1}_{\mathcal{X}} \tag{2.252}$$

is satisfied. When this form of measurement is applied to a register X in a given state $\rho \in \mathrm{D}(\mathcal{X})$, two things happen:

1. An element of Σ is selected at random; each outcome $a \in \Sigma$ is obtained with probability $\langle M_a^* M_a, \rho \rangle$.
2. Conditioned on the measurement outcome $a \in \Sigma$ having been obtained, the state of the register X becomes

$$\frac{M_a \rho M_a^*}{\langle M_a^* M_a, \rho \rangle}. \tag{2.253}$$

Measurements of this sort will be referred to as *nondestructive measurements* in this book.

A somewhat more general notion of a measurement is described by a function of the form $\nu : \Sigma \to \mathrm{CP}(\mathcal{X}, \mathcal{Y})$, again often expressed as a collection

$$\{\Phi_a : a \in \Sigma\} \subset \mathrm{CP}(\mathcal{X}, \mathcal{Y}), \tag{2.254}$$

where Σ is the measurement outcome alphabet, $\mathcal{X}$ is the complex Euclidean space corresponding to the register that is measured, and $\mathcal{Y}$ is an arbitrary complex Euclidean space. In this case, these mappings must necessarily sum to a channel:

$$\sum_{a \in \Sigma} \Phi_a \in \mathrm{C}(\mathcal{X}, \mathcal{Y}). \tag{2.255}$$

Along similar lines to nondestructive measurements, when this form of measurement is applied to a register X in a given state $\rho \in \mathrm{D}(\mathcal{X})$, two things happen:

1. An element of Σ is selected at random, with each outcome $a \in \Sigma$ being obtained with probability $\mathrm{Tr}(\Phi_a(\rho))$.
2. Conditioned on the outcome $a \in \Sigma$ having been obtained, X is transformed into a new register Y having state

$$\frac{\Phi_a(\rho)}{\mathrm{Tr}(\Phi_a(\rho))}. \tag{2.256}$$

The generalized notion of a measurement obtained in this way is called an *instrument*. Nondestructive measurements of the form (2.251) may be represented by instruments of the form (2.254) by defining

$$\Phi_a(X) = M_a X M_a^* \tag{2.257}$$

for each $a \in \Sigma$.

Processes that are expressible as instruments, including nondestructive measurements, can alternatively be described as compositions of channels and (ordinary) measurements. Specifically, for a given instrument of the form (2.254), one may introduce a (classical) register Z having classical state set Σ, and define a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Z} \otimes \mathcal{Y})$ as

$$\Phi(X) = \sum_{a \in \Sigma} E_{a,a} \otimes \Phi_a(X) \tag{2.258}$$

for every $X \in \mathrm{L}(\mathcal{X})$. The fact that Φ is indeed a channel follows directly from the constraints placed on a function of the form (2.254) that must be satisfied for it to be considered an instrument: the complete positivity of the collection of mappings $\{\Phi_a : a \in \Sigma\}$ implies that Φ is completely positive, while the condition (2.255) implies that Φ preserves trace.

Now, if such a channel Φ is applied to a register X, and then the register Z is measured with respect to the standard basis of $\mathcal{Z}$, the distribution of measurement outcomes, as well as the corresponding state of Y conditioned on each possible outcome, is identical to the process associated with the instrument (2.254), as described above.

2.3.3 Extremal measurements and ensembles

Measurements and ensembles may be regarded as elements of convex sets in a fairly straightforward way. A characterization of the extreme points of these sets is obtained below.

Convex combinations of measurements

For $\mathcal{X}$ being a complex Euclidean space and Σ being an alphabet, one may take convex combinations of measurements of the form $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ in the following way. For an alphabet Γ, a probability vector $p \in \mathcal{P}(\Gamma)$, and a collection $\{\mu_b : b \in \Gamma\}$ of measurements taking the form $\mu_b : \Sigma \to \mathrm{Pos}(\mathcal{X})$ for each $b \in \Gamma$, one defines the measurement

$$\mu = \sum_{b \in \Gamma} p(b) \mu_b \tag{2.259}$$

by the equation

$$\mu(a) = \sum_{b \in \Gamma} p(b) \mu_b(a) \tag{2.260}$$

holding for all $a \in \Sigma$. Equivalently, such a convex combination is taken with respect to the most straightforward way of regarding the set of all functions of the form $\theta : \Sigma \to \mathrm{Herm}(\mathcal{X})$ as a vector space over the real numbers.

Another equivalent description of this notion is obtained through the identification of each measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ with its corresponding quantum-to-classical channel

$$\Phi_\mu(X) = \sum_{a \in \Sigma} \langle \mu(a), X \rangle E_{a,a}. \tag{2.261}$$

Convex combinations of measurements then correspond to ordinary convex combinations of their associated channels.

The measurement described by the convex combination (2.259) may be viewed as being equivalent to a process whereby $b \in \Gamma$ is chosen according to the probability vector p, and the measurement μ_b is performed for the chosen symbol $b \in \Gamma$. The outcome of the measurement μ_b is taken as the output of the new measurement, while the symbol $b \in \Gamma$ is discarded.

Extremal measurements

As was established by Proposition 2.38, the set of all quantum-to-classical channels is compact and convex. A measurement is said to be an *extremal measurement* if and only if its corresponding quantum-to-classical channel corresponds to an extreme point of this set.

The definition below states this condition in more concrete terms. A characterization of extremal measurements is provided by the theorem that follows.

Definition 2.46. Let Σ be an alphabet and let $\mathcal{X}$ be a complex Euclidean space. A measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ is an *extremal measurement* if and only if, for all choices of measurements $\mu_0, \mu_1 : \Sigma \to \mathrm{Pos}(\mathcal{X})$ satisfying $\mu = \lambda\mu_0 + (1-\lambda)\mu_1$ for some real number $\lambda \in (0,1)$, one has $\mu_0 = \mu_1$.

Theorem 2.47. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a measurement. It holds that μ is an extremal measurement if and only if, for every function $\theta : \Sigma \to \mathrm{Herm}(\mathcal{X})$ satisfying*

$$\sum_{a \in \Sigma} \theta(a) = 0 \tag{2.262}$$

and $\mathrm{im}(\theta(a)) \subseteq \mathrm{im}(\mu(a))$ for every $a \in \Sigma$, one necessarily has that θ is identically zero: $\theta(a) = 0$ for each $a \in \Sigma$.

Proof. The theorem will be proved in the contrapositive form. Assume first that μ is not an extremal measurement, so that there exist distinct measurements $\mu_0, \mu_1 : \Sigma \to \mathrm{Pos}(\mathcal{X})$ and a scalar value $\lambda \in (0,1)$ for which

$$\mu = \lambda\mu_0 + (1-\lambda)\mu_1. \tag{2.263}$$

It follows that distinct measurements $\nu_0, \nu_1 : \Sigma \to \mathrm{Pos}(\mathcal{X})$ exist for which

$$\mu = \frac{\nu_0 + \nu_1}{2}. \tag{2.264}$$

A suitable choice for these measurements is

$$\begin{aligned} &\nu_0 = 2\lambda\mu_0 + (1-2\lambda)\mu_1 \quad \text{and} \quad \nu_1 = \mu_1, && \text{if } \lambda \leq 1/2; \\ &\nu_0 = \mu_0 \quad \text{and} \quad \nu_1 = (2\lambda-1)\mu_0 + (2-2\lambda)\mu_1, && \text{if } \lambda \geq 1/2. \end{aligned} \tag{2.265}$$

Define $\theta : \Sigma \to \mathrm{Herm}(\mathcal{X})$ as $\theta(a) = \nu_0(a) - \nu_1(a)$ for each $a \in \Sigma$. It holds that

$$\sum_{a\in\Sigma} \theta(a) = \sum_{a\in\Sigma} \nu_0(a) - \sum_{a\in\Sigma} \nu_1(a) = \mathbb{1}_{\mathcal{X}} - \mathbb{1}_{\mathcal{X}} = 0. \tag{2.266}$$

Moreover,

$$\mathrm{im}(\theta(a)) \subseteq \mathrm{im}(\nu_0(a)) + \mathrm{im}(\nu_1(a)) = \mathrm{im}(\mu(a)) \tag{2.267}$$

for each $a \in \Sigma$, where the equality is a consequence of the facts that $\nu_0(a)$ and $\nu_1(a)$ are positive semidefinite and $\mu(a) = (\nu_0(a) + \nu_1(a))/2$. Finally, given that ν_0 and ν_1 are distinct, it is not the case that θ is identically zero.

Now assume that $\theta : \Sigma \to \mathrm{Herm}(\mathcal{X})$ is a function that is not identically zero, and that satisfies

$$\sum_{a\in\Sigma} \theta(a) = 0 \tag{2.268}$$

and $\mathrm{im}(\theta(a)) \subseteq \mathrm{im}(\mu(a))$ for every $a \in \Sigma$. For each $a \in \Sigma$, there must exist a positive real number $\varepsilon_a > 0$ for which

$$\mu(a) + \varepsilon_a\theta(a) \geq 0 \quad \text{and} \quad \mu(a) - \varepsilon_a\theta(a) \geq 0, \tag{2.269}$$

by virtue of the fact that $\mu(a)$ is positive semidefinite and $\theta(a)$ is a Hermitian operator with $\mathrm{im}(\theta(a)) \subseteq \mathrm{im}(\mu(a))$. Let $\varepsilon = \min\{\varepsilon_a : a \in \Sigma\}$ and define

$$\mu_0 = \mu - \varepsilon\theta \quad \text{and} \quad \mu_1 = \mu + \varepsilon\theta. \tag{2.270}$$

It is evident that $\mu = (\mu_0 + \mu_1)/2$. As θ is not identically zero and ε is positive, it holds that μ_0 and μ_1 are distinct. Finally, it holds that μ_0 and μ_1 are measurements; the assumption (2.268) implies that

$$\sum_{a\in\Sigma} \mu_0(a) = \sum_{a\in\Sigma} \mu_1(a) = \sum_{a\in\Sigma} \mu(a) = \mathbb{1}_{\mathcal{X}} \tag{2.271}$$

while the inequalities (2.269) imply that the measurement operators $\mu_0(a)$ and $\mu_1(a)$ are positive semidefinite for each $a \in \Sigma$. It has therefore been established that μ is not an extremal measurement, which completes the proof. □

Theorem 2.47 has various implications, including the corollaries below. The first corollary makes the observation that extremal measurements can have at most $\dim(\mathcal{X})^2$ nonzero measurement operators.

Corollary 2.48. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a measurement. If μ is an extremal measurement, then*

$$\big|\{a \in \Sigma : \mu(a) \neq 0\}\big| \leq \dim(\mathcal{X})^2. \tag{2.272}$$

Proof. The corollary will be proved in the contrapositive form. Let

$$\Gamma = \{a \in \Sigma : \mu(a) \neq 0\}, \tag{2.273}$$

assume that $|\Gamma| > \dim(\mathcal{X})^2$, and consider the collection of measurement operators $\{\mu(a) : a \in \Gamma\}$ as a subset of the real vector space $\mathrm{Herm}(\mathcal{X})$. By the assumption $|\Gamma| > \dim(\mathcal{X})^2$, it must hold that the set $\{\mu(a) : a \in \Gamma\}$ is linearly dependent, and therefore there exist real numbers $\{\alpha_a : a \in \Gamma\}$, not all of which are zero, so that

$$\sum_{a\in\Gamma} \alpha_a \mu(a) = 0. \tag{2.274}$$

Define a function $\theta : \Sigma \to \mathrm{Herm}(\mathcal{X})$ as

$$\theta(a) = \begin{cases} \alpha_a \mu(a) & \text{if } a \in \Gamma \\ 0 & \text{if } a \notin \Gamma. \end{cases} \tag{2.275}$$

It holds that θ is not identically zero, and satisfies

$$\sum_{a\in\Sigma} \theta(a) = 0 \tag{2.276}$$

and $\mathrm{im}(\theta(a)) \subseteq \mathrm{im}(\mu(a))$ for every $a \in \Sigma$. By Theorem 2.47, measurement μ is therefore not an extremal measurement, which completes the proof. □

Corollary 2.48, together with Proposition 2.38 and Theorem 1.10, implies the following corollary.

Corollary 2.49. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a measurement. There exists an alphabet Γ, a probability vector $p \in \mathcal{P}(\Gamma)$, and a collection of measurements $\{\mu_b : b \in \Gamma\}$, taking the form $\mu_b : \Sigma \to \mathrm{Pos}(\mathcal{X})$ and satisfying*

$$\big|\{a \in \Sigma : \mu_b(a) \neq 0\}\big| \leq \dim(\mathcal{X})^2 \tag{2.277}$$

for each $b \in \Gamma$, such that

$$\mu = \sum_{b \in \Gamma} p(b)\mu_b. \tag{2.278}$$

For measurements whose measurement operators all have rank equal to one, Theorem 2.47 yields a simple criterion for extremality, as represented by the following corollary.

Corollary 2.50. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ be a collection of nonzero vectors satisfying*

$$\sum_{a \in \Sigma} x_a x_a^* = \mathbb{1}_{\mathcal{X}}. \tag{2.279}$$

The measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined by $\mu(a) = x_a x_a^*$ for each $a \in \Sigma$ is an extremal measurement if and only if $\{x_a x_a^* : a \in \Sigma\} \subset \mathrm{Herm}(\mathcal{X})$ is a linearly independent set.

Proof. The corollary follows from Theorem 2.47 along with the observation that the set $\{x_a x_a^* : a \in \Sigma\}$ is linearly dependent if and only if there exists a function $\theta : \Sigma \to \mathrm{Herm}(\mathcal{X})$, not identically zero, that satisfies

$$\sum_{a \in \Sigma} \theta(a) = 0 \tag{2.280}$$

and $\mathrm{im}(\theta(a)) \subseteq \mathrm{im}(\mu(a))$ for every $a \in \Sigma$. □

Another implication of Theorem 2.47 is that projective measurements are necessarily extremal.

Corollary 2.51. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a projective measurement. It holds that μ is an extremal measurement.*

Proof. Let $\theta : \Sigma \to \mathrm{Herm}(\mathcal{X})$ be a function satisfying

$$\sum_{a \in \Sigma} \theta(a) = 0 \tag{2.281}$$

and $\mathrm{im}(\theta(a)) \subseteq \mathrm{im}(\mu(a))$ for every $a \in \Sigma$. For each $b \in \Sigma$, it therefore holds that

$$\sum_{a \in \Sigma} \mu(b)\theta(a) = 0. \tag{2.282}$$

By Proposition 2.40 the collection of projections $\{\mu(b) : b \in \Sigma\}$ is an orthogonal collection, and therefore

$$\mu(b)\theta(a) = \begin{cases} \theta(a) & \text{if } a = b \\ 0 & \text{if } a \neq b. \end{cases} \tag{2.283}$$

It follows from (2.282) that $\theta(b) = 0$ for every $b \in \Sigma$, and therefore the function θ is identically zero. As this is so for every choice of θ, as described above, it follows from Theorem 2.47 that μ is an extremal measurement. □

Convex combinations of ensembles of states

Convex combinations of ensembles of states may be defined in essentially the same way that convex combinations of measurements are defined. That is, if $\mathcal{X}$ is a complex Euclidean space, Σ and Γ are alphabets, $p \in \mathcal{P}(\Gamma)$ is a probability vector, and $\eta_b : \Sigma \to \mathrm{Pos}(\mathcal{X})$ is an ensemble of states for each $b \in \Gamma$, then the function $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined by

$$\eta(a) = \sum_{b \in \Gamma} p(b)\eta_b(a) \tag{2.284}$$

for every $a \in \Sigma$ is also an ensemble. One writes

$$\eta = \sum_{b \in \Gamma} p(b)\eta_b \tag{2.285}$$

in this situation. If a density operator $\rho_b \in \mathrm{D}(\mathcal{X})$, representing the average state of the ensemble η_b, is defined as

$$\rho_b = \sum_{a \in \Sigma} \eta_b(a) \tag{2.286}$$

for each $b \in \Gamma$, then it must hold that the average state of the ensemble η is given by

$$\sum_{a \in \Sigma} \eta(a) = \sum_{b \in \Gamma} p(b) \rho_b. \tag{2.287}$$

It is straightforward consequence of the spectral theorem (as represented by Corollary 1.4) that the extreme points of the set of all ensembles of the form $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ take a simple form; they are the ensembles η that are defined as

$$\eta(a) = \begin{cases} uu^* & \text{if } a = b \\ 0 & \text{if } a \neq b, \end{cases} \tag{2.288}$$

for some choice of a unit vector $u \in \mathcal{X}$ and a symbol $b \in \Sigma$.

In some situations, however, it is appropriate to consider just the subset of ensembles of the form $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ that have a particular average state ρ. This set possesses essentially the same convex structure as the set of measurements of the same form. The following proposition establishes one useful fact along these lines.

Proposition 2.52. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble, and let*

$$\rho = \sum_{a \in \Sigma} \eta(a) \tag{2.289}$$

be this ensemble's average state. There exists an alphabet Γ and a collection of ensembles $\{\eta_b : b \in \Gamma\}$ taking the form $\eta_b : \Sigma \to \mathrm{Pos}(\mathcal{X})$ so that the following properties are satisfied:

1. *For each $b \in \Gamma$, the average state of η_b is ρ:*

$$\sum_{a \in \Sigma} \eta_b(a) = \rho. \tag{2.290}$$

2. *For each $b \in \Gamma$, it holds that*

$$\bigl|\{a \in \Sigma : \eta_b(a) \neq 0\}\bigr| \leq \mathrm{rank}(\rho)^2. \tag{2.291}$$

3. *The ensemble η is a convex combination of the ensembles $\{\eta_b : b \in \Gamma\}$. Equivalently, it holds that*

$$\eta = \sum_{b \in \Gamma} p(b) \eta_b \tag{2.292}$$

for some choice of a probability vector $p \in \mathcal{P}(\Gamma)$.

Proof. Let $\mathcal{Y}$ be a complex Euclidean space satisfying $\dim(\mathcal{Y}) = \operatorname{rank}(\rho)$, and let $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be an operator satisfying $AA^* = \rho$. Such an operator A is necessarily invertible. For each $a \in \Sigma$, it holds that

$$\operatorname{im}(\eta(a)) \subseteq \operatorname{im}(\rho) = \operatorname{im}(A). \tag{2.293}$$

By Lemma 2.30, one may therefore conclude that there exists a positive semidefinite operator $Q_a \in \operatorname{Pos}(\mathcal{Y})$ such that

$$\eta(a) = AQ_aA^*, \tag{2.294}$$

for each $a \in \Sigma$.

Now define $\mu : \Sigma \to \operatorname{Pos}(\mathcal{Y})$ as $\mu(a) = Q_a$ for each $a \in \Sigma$. As

$$AA^* = \rho = \sum_{a\in\Sigma} \eta(a) = A\left(\sum_{a\in\Sigma} \mu(a)\right)A^*, \tag{2.295}$$

the invertibility of A implies that

$$\sum_{a\in\Sigma} \mu(a) = \mathbb{1}_{\mathcal{Y}}, \tag{2.296}$$

and therefore μ is a measurement.

By Corollary 2.49, there exists an alphabet Γ, a collection of measurements $\{\mu_b : b \in \Gamma\}$ taking the form $\mu_b : \Sigma \to \operatorname{Pos}(\mathcal{Y})$ and satisfying

$$\big|\{a \in \Sigma : \mu_b(a) \neq 0\}\big| \leq \dim(\mathcal{Y})^2 \tag{2.297}$$

for each $b \in \Gamma$, and a probability vector $p \in \mathcal{P}(\Gamma)$, such that

$$\mu = \sum_{b\in\Gamma} p(b)\mu_b. \tag{2.298}$$

Define a function $\eta_b : \Sigma \to \operatorname{Pos}(\mathcal{X})$ for each $b \in \Gamma$ as

$$\eta_b(a) = A\mu_b(a)A^* \tag{2.299}$$

for each $a \in \Sigma$. It is evident that each η_b is an ensemble whose average state is ρ, by virtue of the fact that each μ_b is a measurement, and the requirement (2.291) follows directly from (2.297). Finally, one has

$$\sum_{b\in\Gamma} p(b)\eta_b(a) = A\left(\sum_{b\in\Gamma} p(b)\mu_b(a)\right)A^* = A\mu(a)A^* = \eta(a) \tag{2.300}$$

for each $a \in \Sigma$, and therefore (2.292) holds, which completes the proof. □

2.4 Exercises

2.1. Let Σ be an alphabet, let $\mathcal{X}$ be a complex Euclidean space, and let

$$\phi : \mathrm{Herm}(\mathcal{X}) \to \mathbb{R}^{\Sigma} \tag{2.301}$$

be a linear function. Prove that these two statements are equivalent:

1. It holds that $\phi(\rho) \in \mathcal{P}(\Sigma)$ for every density operator $\rho \in \mathrm{D}(\mathcal{X})$.
2. There exists a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ such that

$$(\phi(H))(a) = \langle \mu(a), H \rangle \tag{2.302}$$

for every $H \in \mathrm{Herm}(\mathcal{X})$ and $a \in \Sigma$.

2.2. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let Σ be an alphabet, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states. Suppose further that $u \in \mathcal{X} \otimes \mathcal{Y}$ is a vector such that

$$\mathrm{Tr}_{\mathcal{Y}}(uu^*) = \sum_{a \in \Sigma} \eta(a).$$

Prove that there exists a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ for which it holds that

$$\eta(a) = \mathrm{Tr}_{\mathcal{Y}}\big((\mathbb{1}_{\mathcal{X}} \otimes \mu(a))uu^*\big)$$

for all $a \in \Sigma$.

2.3. Let $\Phi \in \mathrm{CP}(\mathcal{X}, \mathcal{Y})$ be a nonzero completely positive map, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces, and let $r = \mathrm{rank}(J(\Phi))$ be the Choi rank of Φ. Prove that there exists a complex Euclidean space $\mathcal{Z}$ having dimension r, along with an operator $A \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Z}, \mathcal{Y})$, such that

$$\Phi(X) = A(X \otimes \mathbb{1}_{\mathcal{Z}})A^* \tag{2.303}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Provide a closed-form equation involving the operator A that is equivalent to Φ preserving trace.

2.4. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a positive map, and let $\Delta \in \mathrm{C}(\mathcal{Y})$ denote the completely dephasing channel with respect to the space $\mathcal{Y}$. Prove that $\Delta\Phi$ is completely positive.

2.5. Let $\Phi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{Z}, \mathcal{Y} \otimes \mathcal{W})$ be a channel, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$. Prove that the following two statements are equivalent:

1. There exists a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ such that
$$\mathrm{Tr}_{\mathcal{W}}\big(J(\Phi)\big) = J(\Psi) \otimes \mathbb{1}_{\mathcal{Z}}. \tag{2.304}$$
2. There exists a complex Euclidean space $\mathcal{V}$ with $\dim(\mathcal{V}) = \dim(\mathcal{X} \otimes \mathcal{Y})$, along with channels $\Phi_0 \in \mathrm{C}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{V})$ and $\Phi_1 \in \mathrm{C}(\mathcal{V} \otimes \mathcal{Z}, \mathcal{W})$, such that
$$\Phi = \big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Phi_1\big)\big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big). \tag{2.305}$$

2.6. Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be complex Euclidean spaces.

(a) Prove that every operator $P \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$ satisfying the equation
$$\big\langle P, J(\Phi)\big\rangle = 1 \tag{2.306}$$
for every channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ must take the form
$$P = \mathbb{1}_{\mathcal{Y}} \otimes \rho \tag{2.307}$$
for some choice of $\rho \in \mathrm{D}(\mathcal{X})$.

(b) Let $\Xi \in \mathrm{CP}(\mathcal{Y} \otimes \mathcal{X}, \mathcal{W} \otimes \mathcal{Z})$ be a completely positive map for which the following statement holds: for every channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, there exists a channel $\Psi \in \mathrm{C}(\mathcal{Z}, \mathcal{W})$ such that
$$\Xi(J(\Phi)) = J(\Psi). \tag{2.308}$$
Prove that there must exist a unital map $\Lambda \in \mathrm{CP}(\mathcal{X}, \mathcal{Z})$ such that
$$\mathrm{Tr}_{\mathcal{W}}\big(\Xi(X)\big) = \Lambda\big(\mathrm{Tr}_{\mathcal{Y}}(X)\big) \tag{2.309}$$
for all $X \in \mathrm{L}(\mathcal{Y} \otimes \mathcal{X})$.

(c) Let $\Xi \in \mathrm{CP}(\mathcal{Y} \otimes \mathcal{X}, \mathcal{W} \otimes \mathcal{Z})$ be a completely positive map satisfying the same requirements as in part (b). Prove that there exists a complex Euclidean space $\mathcal{V}$, along with channels $\Xi_0 \in \mathrm{C}(\mathcal{Z}, \mathcal{X} \otimes \mathcal{V})$ and $\Xi_1 \in \mathrm{C}(\mathcal{Y} \otimes \mathcal{V}, \mathcal{W})$, for which the following property holds: for every channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, the channel $\Psi \in \mathrm{C}(\mathcal{Z}, \mathcal{W})$ that is uniquely determined by (2.308) is given by
$$\Psi = \Xi_1(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{V})})\Xi_0. \tag{2.310}$$

2.5 Bibliographic remarks

The theory of quantum information represents a mathematical formulation of certain aspects of quantum physics, particularly aspects relating to the storage and processing of information in abstract physical systems. While the history of quantum physics is not within the scope of this book, it is appropriate to mention that the mathematical theory discussed in this book is rooted in the work of the many physicists who first developed that field, including Planck, Einstein, Bohr, Heisenberg, Schrödinger, Born, Dirac, and Pauli. Much of this work was placed on a firm mathematical foundation by von Neumann's book *Mathematical Foundations of Quantum Mechanics* [217].

The description of quantum states as density operators was proposed independently by von Neumann [214] and Landau [141] in 1927, a notion equivalent to that of quantum channels was proposed by Haag and Kastler [86] in 1964, and the definition of measurements that has been adopted in this book was proposed by Davies and Lewis [55] in 1970. The relevance of this definition of measurements was articulated by Holevo [104, 106, 107, 108]; in earlier formulations of the theory, only projective measurements were considered. The books of Helstrom [101] and Kraus [135], from 1976 and 1983, respectively, further refined these key foundational aspects of the theory of quantum information.

Further information on the history of quantum information can be found in the books of Peres [169], Nielsen and Chuang [165], and Wilde [229], which are also indispensable references on the theory itself. Kitaev, Shen, and Vyalyi [130] and Bengtsson and Życzkowski [29] also describe the basic formalism that has been presented in this chapter, and include discussions of various specialized topics connected with quantum information.

The Choi representation of maps is so-named for Choi's 1975 paper [50] characterizing completely positive maps (as represented by the equivalence of statements 1 and 3 in Theorem 2.22). Theorem 2.31 was also proved in the same paper. A similar representation to the Choi representation was used earlier by de Pillis [57] and Jamiołkowski [125], and there are arguments to be made for the claim that the representation itself may be considered as folklore.

Theorem 2.22 is an amalgamation of results that are generally attributed to Stinespring [197], Kraus [134, 135], and Choi [50]. Stinespring and Kraus proved more general results for infinite-dimensional spaces; Theorem 2.22

presents only the finite-dimensional analogues of the results they proved. (Several theorems to be presented in this book have a similar character, often having originally been proved in the setting of C*-algebras, as compared with the simpler setting of complex Euclidean spaces.) Theorems 2.25 and 2.26 include equivalences that may be derived from the work of de Pillis [57] and Jamiołkowski [125], respectively.

Theorem 2.42 is a simplified variant of a theorem commonly known as Naimark's theorem (or Naimark's dilation theorem). A more general form of this theorem, holding for certain infinite-dimensional spaces and measure-theoretic formulations of measurements having infinitely many outcomes, was proved by Naimark (whose name is sometimes alternatively transliterated as Neumark) in 1943 [159]. This theorem is now commonly described as being a direct consequence of the later work of Stinespring mentioned above.

The characterization of extremal measurements given by Theorem 2.47 is equivalent to one obtained by Parthasarathy [167]. Results equivalent to Corollaries 2.48, 2.50, and 2.51 were observed in the same paper. The fact that projective measurements are extremal (Corollary 2.51) was also proved earlier by Holevo [108].

Exercise 2.2 is representative of a fact first proved by Hughston, Jozsa, and Wootters [123]. The fact represented by Exercise 2.5 is due to Eggeling, Schlingemann, and Werner [67], answering a question raised by Beckman, Gottesman, Nielsen, and Preskill [25] (who credit DiVincenzo for raising the question). Generalizations of this result to quantum processes having inputs and outputs alternating for multiple steps were obtained by Gutoski and Watrous [85] and Chiribella, D'Ariano, and Perinotti [48]. Exercise 2.6 is representative of a related result of Chiribella, D'Ariano, and Perinotti [47].

Chapter 3

Similarity and distance among states and channels

The main focus of this chapter is on quantifiable notions of similarity and distance between quantum states, the task of discrimination among two or more quantum state alternatives, and related notions involving channels.

There are three main sections of the chapter, the first of which discusses the task of *discrimination* between pairs of quantum states, its connection to the trace norm, and generalizations of this task to more than two states. The second section introduces the *fidelity* function and describes some of its basic properties, formulations, and connections to other concepts. The third section discusses the *completely bounded trace norm*, which is a natural analogue of the trace norm for mappings between spaces of operators, and establishes a connection between this norm and the task of discrimination between pairs of quantum channels.

3.1 Quantum state discrimination

It is a natural question to ask how well a given collection of quantum states can be discriminated by means of a measurement. The hypothetical task of *state discrimination* serves as an abstraction through which this question may be considered.

In the simplest formulation of the state discrimination task, one of two known quantum states is selected at random, and a register prepared in that state is made available to a hypothetical individual. This individual's goal

is to determine which of the two states was selected, by means of a measurement performed on the given register. A theorem known as the *Holevo–Helstrom theorem* gives a closed-form expression, based on the trace norm of a weighted difference between the two possible states, for the probability that an optimally chosen measurement correctly identifies the selected state. An explicit description of an optimal measurement may be obtained from the proof of this theorem.

State discrimination may also be considered in the situation where more than two states are to be discriminated. An analysis of this task is more complicated than the two-state case, and simple, closed-form expressions for the optimal success probability are not known in general. It is possible, however, to represent the optimal success probability through the use of semidefinite programming, which provides a valuable analytical tool through which state discrimination may be analyzed. Approximate solutions, together with bounds on their performance, are also considered.

3.1.1 Discriminating between pairs of quantum states

The simplest formulation of the state discrimination task concerns the discrimination between two fixed quantum states $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ of a given register X. A key aspect of this formulation, together with its analysis, is that it establishes a close connection between the trace norm and the task of state discrimination. Somewhat more generally, one finds that the trace norm provides a natural way of quantifying the "measurable difference" between two quantum states.

Discriminating between pairs of probabilistic states

Before discussing the task of state discrimination between pairs of quantum states, it is appropriate to consider an analogous problem for probabilistic states. To this end, consider the following scenario involving two hypothetical individuals: Alice and Bob.

Scenario 3.1. Let Σ be an alphabet, let X be a classical register with classical state set Σ, and let Y be a classical register with classical state set $\{0, 1\}$. Also let $p_0, p_1 \in \mathcal{P}(\Sigma)$ be probability vectors, representing probabilistic states of X, and let $\lambda \in [0, 1]$ be a real number. The vectors p_0 and p_1, as well as the number λ, are assumed to be known to both Alice and Bob.

Alice prepares Y in a probabilistic state, so that its value is 0 with probability λ and 1 with probability $1 - \lambda$. Conditioned on the value stored in Y, Alice performs one of the following actions:

1. If $\mathsf{Y} = 0$, Alice prepares X in the probabilistic state p_0, and sends X to Bob.
2. If $\mathsf{Y} = 1$, Alice prepares X in the probabilistic state p_1, and sends X to Bob.

Bob's goal is to correctly determine the value of the bit stored in Y, using only the information he can gather from an observation of X.

An optimal strategy in this scenario for Bob, assuming that he wishes to maximize the probability of correctly guessing the value of Y, may be derived from Bayes' theorem, which implies

$$\begin{aligned} \operatorname{Pr}\bigl(\mathsf{Y} = 0 | \mathsf{X} = b\bigr) &= \frac{\lambda p_0(b)}{\lambda p_0(b) + (1-\lambda) p_1(b)} \\ \operatorname{Pr}\bigl(\mathsf{Y} = 1 | \mathsf{X} = b\bigr) &= \frac{(1-\lambda) p_1(b)}{\lambda p_0(b) + (1-\lambda) p_1(b)} \end{aligned} \tag{3.1}$$

for each $b \in \Sigma$. Given the knowledge that $\mathsf{X} = b$, Bob should therefore choose the more likely value for Y: if it holds that $\lambda p_0(b) > (1-\lambda) p_1(b)$, then Bob should guess that $\mathsf{Y} = 0$, while if $\lambda p_0(b) < (1-\lambda) p_1(b)$, then Bob should guess that $\mathsf{Y} = 1$. In the case that $\lambda p_0(b) = (1-\lambda) p_1(b)$, Bob can guess either $\mathsf{Y} = 0$ or $\mathsf{Y} = 1$ arbitrarily without affecting his probability of being correct, as the two values are equally likely in this situation.

The probability that Bob correctly identifies the value of Y using this strategy can be understood by first considering the probability he is correct *minus* the probability he is incorrect. This difference in probabilities is represented by the quantity

$$\sum_{b \in \Sigma} \bigl| \lambda p_0(b) - (1-\lambda) p_1(b) \bigr| = \bigl\| \lambda p_0 - (1-\lambda) p_1 \bigr\|_1. \tag{3.2}$$

It follows that the probability that Bob is correct is given by the quantity

$$\frac{1}{2} + \frac{1}{2} \bigl\| \lambda p_0 - (1-\lambda) p_1 \bigr\|_1. \tag{3.3}$$

This expression makes clear the close connection between probabilistic state discrimination and the vector 1-norm.

Notice that

$$0 \leq \left\| \lambda p_0 - (1-\lambda) p_1 \right\|_1 \leq 1, \tag{3.4}$$

where the second inequality follows from the triangle inequality. This is consistent with the interpretation of the expression (3.3) as a probability. In the extreme case where

$$\left\| \lambda p_0 - (1-\lambda) p_1 \right\|_1 = 0, \tag{3.5}$$

which requires $\lambda = 1/2$ and $p_0 = p_1$, Bob is essentially reduced to guessing blindly and will be correct with probability $1/2$. In the other extreme,

$$\left\| \lambda p_0 - (1-\lambda) p_1 \right\|_1 = 1, \tag{3.6}$$

it must hold that λp_0 and $(1-\lambda) p_1$ have disjoint supports, and thus Bob can determine the value of Y without error. Intermediate values, in which both inequalities in (3.4) hold strictly, correspond to different degrees of certainty in Bob's guess.

Discriminating between pairs of quantum states

The task of discriminating between pairs of quantum states is represented by the following scenario, which is the natural quantum generalization of Scenario 3.1.

Scenario 3.2. Let X be an arbitrary register and let Y be a classical register with classical state set $\{0,1\}$. Also let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be states of X, and let $\lambda \in [0,1]$ be a real number. The states ρ_0 and ρ_1, as well as the number λ, are assumed to be known to both Alice and Bob.

Alice prepares Y in a probabilistic state, so that its value is 0 with probability λ and 1 with probability $1-\lambda$. Conditioned on the value stored in Y, Alice performs one of the following actions:

1. If $\mathsf{Y} = 0$, Alice prepares X in the state ρ_0, and sends X to Bob.
2. If $\mathsf{Y} = 1$, Alice prepares X in the state ρ_1, and sends X to Bob.

Bob's goal is to correctly determine the value of the bit stored in Y, by means of a measurement of X.

The principal goal of the discussion that follows is to establish an analogous connection between this scenario and the trace norm to the one that was shown above to hold between Scenario 3.1 and the vector 1-norm. The following lemma, which happens to concern the spectral norm rather than the trace norm, is useful for establishing this connection. The lemma is stated in greater generality than is required for the purposes of the present section, but the more general form will find uses elsewhere in this book.

Lemma 3.3. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let $u \in \mathbb{C}^{\Sigma}$ be a vector, and let $\{P_a : a \in \Sigma\} \subset \mathrm{Pos}(\mathcal{X})$ be a collection of positive semidefinite operators. It holds that*

$$\left\| \sum_{a \in \Sigma} u(a) P_a \right\| \leq \|u\|_\infty \left\| \sum_{a \in \Sigma} P_a \right\|. \tag{3.7}$$

Proof. Define an operator $A \in \mathrm{L}\big(\mathcal{X}, \mathcal{X} \otimes \mathbb{C}^{\Sigma}\big)$ as

$$A = \sum_{a \in \Sigma} \sqrt{P_a} \otimes e_a. \tag{3.8}$$

The spectral norm is submultiplicative with respect to compositions and multiplicative with respect to tensor products, and therefore

$$\begin{aligned} \left\| \sum_{a \in \Sigma} u(a) P_a \right\| &= \left\| \sum_{a \in \Sigma} u(a) A^* (\mathbb{1}_{\mathcal{X}} \otimes E_{a,a}) A \right\| \\ &\leq \|A^*\| \left\| \sum_{a \in \Sigma} u(a) E_{a,a} \right\| \|A\| = \|u\|_\infty \|A\|^2. \end{aligned} \tag{3.9}$$

Finally, by the spectral norm property (1.172), one has

$$\|A\|^2 = \|A^* A\| = \left\| \sum_{a \in \Sigma} P_a \right\|, \tag{3.10}$$

which completes the proof. □

A direct connection between Scenario 3.2 and the trace norm can now be established. The next theorem, known as the Holevo–Helstrom theorem, expresses this connection in mathematical terms.

Theorem 3.4 (Holevo–Helstrom theorem). *Let $\mathcal{X}$ be a complex Euclidean space, let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be density operators, and let $\lambda \in [0,1]$. For every choice of a measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$, it holds that*

$$\lambda \langle \mu(0), \rho_0 \rangle + (1-\lambda)\langle \mu(1), \rho_1 \rangle \leq \frac{1}{2} + \frac{1}{2} \big\| \lambda \rho_0 - (1-\lambda)\rho_1 \big\|_1. \tag{3.11}$$

Moreover, there exists a projective measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$ for which equality is achieved in (3.11).

Proof. Define

$$\rho = \lambda \rho_0 + (1-\lambda)\rho_1 \quad \text{and} \quad X = \lambda \rho_0 - (1-\lambda)\rho_1, \tag{3.12}$$

so that

$$\lambda \rho_0 = \frac{\rho + X}{2} \quad \text{and} \quad (1-\lambda)\rho_1 = \frac{\rho - X}{2}, \tag{3.13}$$

and therefore

$$\lambda \langle \mu(0), \rho_0 \rangle + (1-\lambda)\langle \mu(1), \rho_1 \rangle = \frac{1}{2} + \frac{1}{2} \langle \mu(0) - \mu(1), X \rangle. \tag{3.14}$$

By Lemma 3.3, together with the Hölder inequality for Schatten norms, it follows that

$$\begin{aligned} &\frac{1}{2} + \frac{1}{2} \langle \mu(0) - \mu(1), X \rangle \\ &\quad \leq \frac{1}{2} + \frac{1}{2} \big\| \mu(0) - \mu(1) \big\| \, \big\| X \big\|_1 \leq \frac{1}{2} + \frac{1}{2} \big\| X \big\|_1. \end{aligned} \tag{3.15}$$

Combining (3.14) and (3.15) yields (3.11).

To show that equality is achieved in (3.11) for a projective measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$, one may consider the Jordan-Hahn decomposition

$$X = P - Q, \tag{3.16}$$

for $P, Q \in \mathrm{Pos}(\mathcal{X})$. Define $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$ as

$$\mu(0) = \Pi_{\mathrm{im}(P)} \quad \text{and} \quad \mu(1) = \mathbb{1} - \Pi_{\mathrm{im}(P)}, \tag{3.17}$$

which is a projective measurement. It holds that

$$\langle \mu(0) - \mu(1), X \rangle = \mathrm{Tr}(P) + \mathrm{Tr}(Q) = \big\| X \big\|_1, \tag{3.18}$$

and therefore

$$\lambda \langle \mu(0), \rho_0 \rangle + (1-\lambda)\langle \mu(1), \rho_1 \rangle = \frac{1}{2} + \frac{1}{2} \big\| X \big\|_1, \tag{3.19}$$

which completes the proof. □

It follows from Theorem 3.4 that an optimal choice of a measurement for Bob in Scenario 3.2 correctly determines the value of Y with probability

$$\frac{1}{2} + \frac{1}{2}\left\| \lambda\rho_0 - (1-\lambda)\rho_1 \right\|_1, \tag{3.20}$$

and, moreover, this optimal probability is achievable with a projective measurement.

One might question the implicit claim that the possible strategies for Bob in Scenario 3.2 are exhausted by the consideration of measurements having 0 and 1 as the only possible outcomes. For instance, Bob could measure X using a measurement with three or more outcomes, and then base his guess for the value of Y on some sort of post-processing of the measurement outcome obtained. However, no generality is introduced by this type of strategy, or any other strategy having access to the register X alone. Any process used by Bob to eventually produce a binary-valued guess for the classical state of Y must define a binary-valued measurement, and Theorem 3.4 may be applied to this measurement.

The following proposition, whose proof has some overlap with the proof of the Theorem 3.4, establishes a useful relationship between the trace norm of an operator and the 1-norm of a vector obtained from that operator's inner products with the measurement operators of any measurement.

Proposition 3.5. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a measurement, and let $X \in \mathrm{L}(\mathcal{X})$ be an operator. Define a vector $v \in \mathbb{C}^{\Sigma}$ as*

$$v(a) = \langle \mu(a), X \rangle \tag{3.21}$$

for each $a \in \Sigma$. It holds that $\|v\|_1 \leq \|X\|_1$.

Proof. One has

$$\|v\|_1 = \sum_{a\in\Sigma} \left| \langle \mu(a), X \rangle \right| = \sum_{a\in\Sigma} u(a) \langle \mu(a), X \rangle = \left\langle \sum_{a\in\Sigma} \overline{u(a)}\mu(a), X \right\rangle \tag{3.22}$$

for some choice of a vector $u \in \mathbb{C}^{\Sigma}$ satisfying $|u(a)| = 1$ for each $a \in \Sigma$. By Lemma 3.3, together with Hölder's inequality for Schatten norms, it follows that

$$\|v\|_1 \leq \left\| \sum_{a\in\Sigma} \overline{u(a)}\mu(a) \right\| \|X\|_1 \leq \|X\|_1, \tag{3.23}$$

as required. □

Discriminating between convex sets of quantum states

The task of state discrimination between pairs of quantum states may be generalized to one in which two convex sets of quantum states are to be discriminated. The following scenario describes this task in more precise terms.

Scenario 3.6. Let X be an arbitrary register and let Y be a classical register having classical state set $\{0,1\}$. Also let $\mathcal{C}_0, \mathcal{C}_1 \subseteq \mathrm{D}(\mathcal{X})$ be nonempty, convex sets of states, and let $\lambda \in [0,1]$ be a real number. The sets $\mathcal{C}_0$ and $\mathcal{C}_1$, as well as the number λ, are assumed to be known to both Alice and Bob.

Alice prepares Y in a probabilistic state, so that its value is 0 with probability λ and 1 with probability $1-\lambda$. Conditioned on the value stored in Y, Alice performs one of the following actions:

1. If $\mathsf{Y} = 0$, Alice prepares X in any state $\rho_0 \in \mathcal{C}_0$ of her choice, and sends X to Bob.
2. If $\mathsf{Y} = 1$, Alice prepares X in any state $\rho_1 \in \mathcal{C}_1$ of her choice, and sends X to Bob.

Bob's goal is to predict the value of the bit stored in Y, by means of a measurement of X.

The description of Scenario 3.6 does not specify how Alice is to choose ρ_0 or ρ_1, beyond stating the requirement that $\rho_0 \in \mathcal{C}_0$ and $\rho_1 \in \mathcal{C}_1$. It could be, for instance, that Alice chooses these states randomly according to fixed distributions, or she could choose the states adversarially, even based on a knowledge of the measurement Bob intends to use. What is relevant is that Bob can make no assumptions regarding Alice's choices for ρ_0 and ρ_1, beyond the requirement that she chooses $\rho_0 \in \mathcal{C}_0$ and $\rho_1 \in \mathcal{C}_1$.

One may note that Scenario 3.2 represents a special case of Scenario 3.6 in which $\mathcal{C}_0$ and $\mathcal{C}_1$ are the singleton sets $\{\rho_0\}$ and $\{\rho_1\}$, respectively.

It follows from the Holevo–Helstrom theorem (Theorem 3.4) that Bob cannot hope to succeed in his task in Scenario 3.6 with probability higher than

$$\frac{1}{2} + \frac{1}{2} \bigl\| \lambda \rho_0 - (1-\lambda)\rho_1 \bigr\|_1 , \tag{3.24}$$

for whichever states $\rho_0 \in \mathcal{C}_0$ and $\rho_1 \in \mathcal{C}_1$ Alice chooses, for this is his optimal success probability when he has the additional knowledge that Alice

chooses either ρ_0 or ρ_1. The following proposition implies that Bob can succeed with probability at least

$$\frac{1}{2} + \frac{1}{2} \inf_{\rho_0,\rho_1} \left\| \lambda\rho_0 - (1-\lambda)\rho_1 \right\|_1, \tag{3.25}$$

where the infimum is taken over all choices of $\rho_0 \in \mathcal{C}_0$ and $\rho_1 \in \mathcal{C}_1$. In light of the limitation imposed by the Holevo–Helstrom theorem, this is necessarily the optimal probability of success in the worst case.

Theorem 3.7. *Let $\mathcal{C}_0, \mathcal{C}_1 \subseteq \mathrm{D}(\mathcal{X})$ be nonempty, convex sets, for $\mathcal{X}$ being a complex Euclidean space, and let $\lambda \in [0,1]$. It holds that*

$$\begin{aligned} \max_{\mu} \inf_{\rho_0,\rho_1} \Big(\lambda \langle \mu(0), \rho_0 \rangle + (1-\lambda) \langle \mu(1), \rho_1 \rangle \Big) & \\ = \inf_{\rho_0,\rho_1} \max_{\mu} \Big(\lambda \langle \mu(0), \rho_0 \rangle + (1-\lambda) \langle \mu(1), \rho_1 \rangle \Big) & \\ = \frac{1}{2} + \frac{1}{2} \inf_{\rho_0,\rho_1} \left\| \lambda\rho_0 - (1-\lambda)\rho_1 \right\|_1, & \end{aligned} \tag{3.26}$$

where the infima are over all choices of $\rho_0 \in \mathcal{C}_0$ and $\rho_1 \in \mathcal{C}_1$, and the maxima are over all choices of binary measurements $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$.

Proof. Define sets $\mathcal{A}, \mathcal{B} \subset \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X})$ as

$$\mathcal{A} = \left\{ \begin{pmatrix} \rho_0 & 0 \\ 0 & \rho_1 \end{pmatrix} : \rho_0 \in \mathcal{C}_0,\, \rho_1 \in \mathcal{C}_1 \right\} \tag{3.27}$$

and

$$\mathcal{B} = \left\{ \begin{pmatrix} \lambda P_0 & 0 \\ 0 & (1-\lambda) P_1 \end{pmatrix} : P_0, P_1 \in \mathrm{Pos}(\mathcal{X}),\, P_0 + P_1 = \mathbb{1}_{\mathcal{X}} \right\}, \tag{3.28}$$

as well as a function $f : \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ as $f(A,B) = \langle A, B \rangle$. It holds that $\mathcal{A}$ and $\mathcal{B}$ are convex, $\mathcal{B}$ is compact, and f is bilinear, so that

$$\inf_{A \in \mathcal{A}} \max_{B \in \mathcal{B}} f(A,B) = \max_{B \in \mathcal{B}} \inf_{A \in \mathcal{A}} f(A,B) \tag{3.29}$$

holds by Sion's min-max theorem (Theorem 1.12). The equation (3.29) is equivalent to the first equality of (3.26), and the second equality in (3.26) follows from Theorem 3.4. □

3.1.2 Discriminating quantum states of an ensemble

The remaining variant of quantum state discrimination to be discussed in this chapter is similar to the one represented by Scenario 3.2, except that more than two possible states, selected from a given ensemble, are to be discriminated. The following scenario describes this task in more precise terms.

Scenario 3.8. Let X be an arbitrary register, let Σ be an alphabet, let Y be a classical register having classical state set Σ, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states. Alice prepares the pair (Y, X) in the classical-quantum state

$$\sigma = \sum_{a \in \Sigma} E_{a,a} \otimes \eta(a) \tag{3.30}$$

determined by the ensemble η. Equivalently, the register Y takes each value $a \in \Sigma$ with probability $p(a) = \mathrm{Tr}(\eta(a))$, and conditioned on the event $\mathsf{Y} = a$ the state of X is set to $\eta(a)/\mathrm{Tr}(\eta(a))$, for each $a \in \Sigma$. The register X is sent to Bob, and Bob's goal is to predict the classical state of Y, using only the information he can gather from a measurement of X.

For any measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ chosen by Bob in this scenario, the probability that he correctly predicts the classical state of Y is given by the expression

$$\sum_{a \in \Sigma} \langle \mu(a), \eta(a) \rangle. \tag{3.31}$$

It is therefore natural to consider a maximization of this quantity over all choices of the measurement μ.

More generally, one may substitute an arbitrary function of the form $\phi : \Sigma \to \mathrm{Herm}(\mathcal{X})$ in place of the ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$, and consider a maximization of the quantity

$$\sum_{a \in \Sigma} \langle \mu(a), \phi(a) \rangle \tag{3.32}$$

over all measurements $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$. One situation in which this more general optimization problem is meaningful is a variant of Scenario 3.8 in which different payoff values are associated to each pair (a, b), representing the state a of Alice's register Y and Bob's measurement outcome b. If Bob receives a payoff value of $K(a, b)$ for producing the measurement outcome

b when Alice's register Y holds the symbol a, for instance, Bob's expected gain for a given measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ is given by

$$\sum_{a\in\Sigma}\sum_{b\in\Sigma} K(a,b)\langle \mu(b), \eta(a)\rangle = \sum_{b\in\Sigma}\langle \mu(b), \phi(b)\rangle \tag{3.33}$$

for

$$\phi(b) = \sum_{a\in\Sigma} K(a,b)\,\eta(a). \tag{3.34}$$

This sort of hypothetical situation could be further generalized by allowing the classical state set of Alice's register Y and Bob's set of measurement outcomes to disagree.

A semidefinite program for optimal measurements

For any choice of a complex Euclidean space $\mathcal{X}$, an alphabet Σ, and a function $\phi : \Sigma \to \mathrm{Herm}(\mathcal{X})$, define

$$\mathrm{opt}(\phi) = \max_{\mu} \sum_{a\in\Sigma} \langle \mu(a), \phi(a)\rangle, \tag{3.35}$$

where the maximum is over all measurements $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$. This optimal value is necessarily achieved for some choice of a measurement, as it is a maximization of a continuous function over the compact set of measurements of the form $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, which justifies the use of the maximum rather than the supremum. It may also be said that a particular choice of a measurement μ is *optimal* for ϕ if the above expression (3.32) coincides with the value $\mathrm{opt}(\phi)$.

There is no closed-form expression that is known to represent the value $\mathrm{opt}(\phi)$ for an arbitrary choice of a function $\phi : \Sigma \to \mathrm{Herm}(\mathcal{X})$. However, it is possible to express the value $\mathrm{opt}(\phi)$ by a semidefinite program, providing a method by which it may be numerically calculated using a computer. A simplified description of the primal and dual problems associated with such a semidefinite program are as follows:

Primal problem (simplified)		Dual problem (simplified)	
maximize:	$\sum_{a\in\Sigma}\langle \mu(a), \phi(a)\rangle$	minimize:	$\mathrm{Tr}(Y)$
subject to:	$\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$,	subject to:	$Y \geq \phi(a)$ (for all $a \in \Sigma$),
	$\sum_{a\in\Sigma} \mu(a) = \mathbb{1}_{\mathcal{X}}$.		$Y \in \mathrm{Herm}(\mathcal{X})$.

A formal expression of this semidefinite program that conforms to the definition of semidefinite programs presented in Section 1.2.2 is given by the triple $(\Phi, A, \mathbb{1}_{\mathcal{X}})$, where the mapping $\Phi \in \mathrm{T}(\mathcal{Y} \otimes \mathcal{X}, \mathcal{X})$ is defined as the partial trace $\Phi = \mathrm{Tr}_{\mathcal{Y}}$, for $\mathcal{Y} = \mathbb{C}^{\Sigma}$, and the operator A is defined as

$$A = \sum_{a \in \Sigma} E_{a,a} \otimes \phi(a). \tag{3.36}$$

The primal and dual problems associated with the triple $(\Phi, A, \mathbb{1}_{\mathcal{X}})$ are as follows:

Primal problem (formal)		Dual problem (formal)	
maximize:	$\langle A, X \rangle$	minimize:	$\mathrm{Tr}(Y)$
subject to:	$\mathrm{Tr}_{\mathcal{Y}}(X) = \mathbb{1}_{\mathcal{X}}$,	subject to:	$\mathbb{1}_{\mathcal{Y}} \otimes Y \geq A$,
	$X \in \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{X})$.		$Y \in \mathrm{Herm}(\mathcal{X})$.

These problems are equivalent to the simplified primal and dual problems described above. In greater detail, any feasible solution μ to the simplified primal problem described above gives rise to the feasible solution

$$X = \sum_{a \in \Sigma} E_{a,a} \otimes \mu(a) \tag{3.37}$$

to the formal primal problem, in which the same objective value

$$\langle A, X \rangle = \sum_{a \in \Sigma} \langle \mu(a), \phi(a) \rangle \tag{3.38}$$

is achieved. While a feasible solution X to the formal primal problem need not take the form (3.37) in general, one may nevertheless obtain a feasible solution μ to the simplified primal problem from such an operator X by setting

$$\mu(a) = \big(e_a^* \otimes \mathbb{1}_{\mathcal{X}}\big) X \big(e_a \otimes \mathbb{1}_{\mathcal{X}}\big) \tag{3.39}$$

for each $a \in \Sigma$. The equality (3.38) again holds, and therefore the two primal problems have the same optimal values. The fact that the two dual problems are equivalent is evident from the observation that the inequality

$$\mathbb{1}_{\mathcal{Y}} \otimes Y \geq \sum_{a \in \Sigma} E_{a,a} \otimes \phi(a) \tag{3.40}$$

is equivalent to the inequality $Y \geq \phi(a)$ holding for every $a \in \Sigma$.

Strong duality holds for this semidefinite program. The operator

$$X = \frac{1}{|\Sigma|} \mathbb{1}_{\mathcal{Y}} \otimes \mathbb{1}_{\mathcal{X}} \tag{3.41}$$

is a strictly feasible primal solution, while $Y = \gamma \mathbb{1}_{\mathcal{X}}$ is a strictly feasible dual solution for any real value $\gamma > \lambda_1(A)$. It follows from Slater's theorem (Theorem 1.13) that the optimal primal and dual values for the semidefinite program are equal, and moreover the optimum value is achieved in both the primal and dual problems.

Criteria for measurement optimality

It may be difficult to obtain an analytic description of a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ that is optimal for a given function $\phi : \Sigma \to \mathrm{Herm}(\mathcal{X})$, given the lack of a known closed-form expression for such a measurement. In contrast, it is straightforward to verify that an optimal measurement is indeed optimal by means of the following theorem.

Theorem 3.9 (Holevo–Yuen–Kennedy–Lax)**.** *Let $\phi : \Sigma \to \mathrm{Herm}(\mathcal{X})$ be a function and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be a measurement, for $\mathcal{X}$ being a complex Euclidean space and Σ being an alphabet. It holds that*

$$\sum_{a \in \Sigma} \langle \mu(a), \phi(a) \rangle = \mathrm{opt}(\phi) \tag{3.42}$$

(i.e., the measurement μ is optimal for the function ϕ) if and only if the operator

$$\sum_{a \in \Sigma} \phi(a)\mu(a) \tag{3.43}$$

is Hermitian and satisfies

$$\sum_{a \in \Sigma} \phi(a)\mu(a) \geq \phi(b) \tag{3.44}$$

for every $b \in \Sigma$.

Proof. Define an operator $X \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})$ as

$$X = \sum_{a \in \Sigma} E_{a,a} \otimes \mu(a). \tag{3.45}$$

Suppose first that μ is an optimal measurement for ϕ, so that X is an optimal primal solution to the semidefinite program $(\Phi, A, \mathbb{1}_{\mathcal{X}})$ representing $\operatorname{opt}(\phi)$, as described previously. As the dual optimum of this semidefinite program is always achieved, one may choose $Y \in \operatorname{Herm}(\mathcal{X})$ to be such a dual-optimal solution. By the property of complementary slackness for semidefinite programs (Proposition 1.14), it necessarily holds that

$$(\mathbb{1}_{\mathcal{Y}} \otimes Y)X = AX. \tag{3.46}$$

Taking the partial trace of both sides of (3.46) over $\mathcal{Y}$, one finds that

$$Y = Y \operatorname{Tr}_{\mathcal{Y}}(X) = \operatorname{Tr}_{\mathcal{Y}}(AX) = \sum_{a \in \Sigma} \phi(a)\mu(a). \tag{3.47}$$

The dual-feasibility of Y therefore implies that the operator (3.43) is Hermitian and satisfies (3.44).

To prove the reverse implication, note that if the operator (3.43) is Hermitian and satisfies (3.44) for every $b \in \Sigma$, then

$$Y = \sum_{a \in \Sigma} \phi(a)\mu(a) \tag{3.48}$$

is a dual-feasible solution to the semidefinite program $(\Phi, A, \mathbb{1}_{\mathcal{X}})$ representing $\operatorname{opt}(\phi)$. The operator X defined in (3.45) is a primal-feasible solution to this semidefinite program, simply by virtue of the fact that μ is a measurement. The objective values achieved by X in the primal problem and Y in the dual problem are both equal to

$$\sum_{a \in \Sigma} \langle \mu(a), \phi(a) \rangle. \tag{3.49}$$

The equality between these values implies that both are optimal by the property of weak duality of semidefinite programs. The measurement μ is therefore optimal for ϕ. □

The pretty good measurement

Returning to Bob's task, as described in Scenario 3.8, suppose an ensemble $\eta : \Sigma \to \operatorname{Pos}(\mathcal{X})$ is given, and a measurement $\mu : \Sigma \to \operatorname{Pos}(\mathcal{X})$ maximizing the probability

$$\sum_{a \in \Sigma} \langle \mu(a), \eta(a) \rangle \tag{3.50}$$

of a correct determination of the state of Alice's classical register Y is sought.

In a concrete setting in which an explicit description of η is known, the semidefinite programming formulation of $\mathrm{opt}(\eta)$ allows for an efficient numerical approximation to a measurement μ that is optimal for η. This approach may, however, be unsatisfactory in more abstract settings, such as ones in which it is necessary to view η as being indeterminate. Although Theorem 3.9 allows for a verification that a given optimal measurement is indeed optimal, it does not provide a method to find a measurement that is optimal.

One alternative to searching for an optimal measurement is to consider measurements that are determined from η by closed-form expressions, but that might be sub-optimal. The so-called *pretty good measurement* is an example of such a measurement.

To define the pretty good measurement for a given ensemble η, one first considers the average state

$$\rho = \sum_{a \in \Sigma} \eta(a) \tag{3.51}$$

of η. In the case that ρ is positive definite, the pretty good measurement associated with η is the measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined as

$$\mu(a) = \rho^{-\frac{1}{2}} \eta(a) \rho^{-\frac{1}{2}}. \tag{3.52}$$

In general, when ρ is not necessarily invertible, one may use the Moore–Penrose pseudo-inverse of ρ, in place of the inverse of ρ, to define[1] the pretty good measurement associated with η as

$$\mu(a) = \sqrt{\rho^+}\, \eta(a) \sqrt{\rho^+} + \frac{1}{|\Sigma|} \Pi_{\ker(\rho)} \tag{3.54}$$

for every $a \in \Sigma$.

The pretty good measurement will generally not be optimal for a given ensemble. It does, however, achieve a probability of a correct prediction that

[1] It should be noted that, although the equation (3.54) is taken here as the definition of the pretty good measurement, it is somewhat arbitrary in the case that ρ is not invertible. Any measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ satisfying

$$\mu(a) \geq \sqrt{\rho^+}\, \eta(a) \sqrt{\rho^+} \tag{3.53}$$

for all $a \in \Sigma$ would be equivalent with respect to the discussion that follows.

is at least the square of the optimal success probability, as the following theorem states.

Theorem 3.10 (Barnum–Knill)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states, and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ denote the pretty good measurement associated with η. It holds that*

$$\sum_{a\in\Sigma} \langle \mu(a), \eta(a)\rangle \geq \mathrm{opt}(\eta)^2. \tag{3.55}$$

Proof. Let $\rho = \sum_{a\in\Sigma} \eta(a)$ and let $\nu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be any measurement. For every $a \in \Sigma$ it holds that $\mathrm{im}(\eta(a)) \subseteq \mathrm{im}(\rho)$, and therefore

$$\langle \nu(a), \eta(a)\rangle = \left\langle \rho^{\frac{1}{4}}\nu(a)\rho^{\frac{1}{4}}, \left(\rho^{+}\right)^{\frac{1}{4}} \eta(a) \left(\rho^{+}\right)^{\frac{1}{4}} \right\rangle. \tag{3.56}$$

By the Cauchy–Schwarz inequality, it follows that

$$\langle \nu(a), \eta(a)\rangle \leq \left\| \rho^{\frac{1}{4}}\nu(a)\rho^{\frac{1}{4}} \right\|_2 \left\| \left(\rho^{+}\right)^{\frac{1}{4}} \eta(a) \left(\rho^{+}\right)^{\frac{1}{4}} \right\|_2 \tag{3.57}$$

for each $a \in \Sigma$. Applying the Cauchy–Schwarz inequality again, this time for vectors of real numbers rather than for operators, one finds that

$$\sum_{a\in\Sigma} \langle \nu(a), \eta(a)\rangle \leq \sqrt{\sum_{a\in\Sigma} \left\| \rho^{\frac{1}{4}}\nu(a)\rho^{\frac{1}{4}} \right\|_2^2} \sqrt{\sum_{a\in\Sigma} \left\| \left(\rho^{+}\right)^{\frac{1}{4}} \eta(a) \left(\rho^{+}\right)^{\frac{1}{4}} \right\|_2^2}. \tag{3.58}$$

The first term on the right-hand side of (3.58) is at most 1. To verify that this is so, one may first use the definition of the Frobenius norm to obtain the expression

$$\left\| \rho^{\frac{1}{4}}\nu(a)\rho^{\frac{1}{4}} \right\|_2^2 = \left\langle \rho^{\frac{1}{4}}\nu(a)\rho^{\frac{1}{4}}, \rho^{\frac{1}{4}}\nu(a)\rho^{\frac{1}{4}} \right\rangle = \left\langle \nu(a), \sqrt{\rho}\nu(a)\sqrt{\rho} \right\rangle \tag{3.59}$$

for each $a \in \Sigma$, from which it follows that

$$\left\| \rho^{\frac{1}{4}}\nu(a)\rho^{\frac{1}{4}} \right\|_2^2 \leq \mathrm{Tr}\left(\sqrt{\rho}\nu(a)\sqrt{\rho}\right), \tag{3.60}$$

by virtue of the fact that $\nu(a) \leq \mathbb{1}_{\mathcal{X}}$ and $\sqrt{\rho}\nu(a)\sqrt{\rho} \geq 0$. Summing over all $a \in \Sigma$ yields

$$\sum_{a\in\Sigma} \left\| \rho^{\frac{1}{4}}\nu(a)\rho^{\frac{1}{4}} \right\|_2^2 \leq \sum_{a\in\Sigma} \mathrm{Tr}\left(\sqrt{\rho}\nu(a)\sqrt{\rho}\right) = \mathrm{Tr}(\rho) = 1. \tag{3.61}$$

By the definition of the pretty good measurement, along with a similar computation to the one expressed by (3.59), one has that

$$\left\| (\rho^+)^{\frac{1}{4}} \eta(a) (\rho^+)^{\frac{1}{4}} \right\|_2^2 = \left\langle \sqrt{\rho^+}\, \eta(a) \sqrt{\rho^+}, \eta(a) \right\rangle \leq \langle \mu(a), \eta(a) \rangle \tag{3.62}$$

for each $a \in \Sigma$, and therefore

$$\sum_{a \in \Sigma} \left\| (\rho^+)^{\frac{1}{4}} \eta(a) (\rho^+)^{\frac{1}{4}} \right\|_2^2 \leq \sum_{a \in \Sigma} \langle \mu(a), \eta(a) \rangle. \tag{3.63}$$

By (3.58), (3.61), and (3.63) it follows that

$$\left(\sum_{a \in \Sigma} \langle \nu(a), \eta(a) \rangle \right)^2 \leq \sum_{a \in \Sigma} \langle \mu(a), \eta(a) \rangle. \tag{3.64}$$

As this inequality holds for all measurements $\nu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, including those measurements that are optimal for η, the proof is complete. □

3.2 The fidelity function

This section introduces the *fidelity function*, which provides a measure of the similarity, or "overlap," between quantum states (and positive semidefinite operators more generally) that will be used extensively throughout this book. It is defined as follows.

Definition 3.11. Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. The *fidelity* $\mathrm{F}(P, Q)$ between P and Q is defined as

$$\mathrm{F}(P, Q) = \left\| \sqrt{P} \sqrt{Q} \right\|_1. \tag{3.65}$$

The function F is called the *fidelity function*.

The fidelity function is most often considered for density operator inputs, but there is value in defining it more generally, allowing its arguments to range over arbitrary positive semidefinite operators. An alternative expression for the fidelity function is obtained by expanding (3.65) according to the definition of the trace norm:

$$\mathrm{F}(P, Q) = \mathrm{Tr}\left(\sqrt{\sqrt{P} Q \sqrt{P}} \right). \tag{3.66}$$

3.2.1 Elementary properties of the fidelity function

The following proposition establishes several basic properties of the fidelity function.

Proposition 3.12. *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. The following facts hold:*

1. *The fidelity function* F *is continuous at* (P, Q).
2. $\mathrm{F}(P, Q) = \mathrm{F}(Q, P)$.
3. $\mathrm{F}(\lambda P, Q) = \sqrt{\lambda}\,\mathrm{F}(P, Q) = \mathrm{F}(P, \lambda Q)$ *for every nonnegative real number* λ.
4. $\mathrm{F}(P, Q) = \mathrm{F}(P, \Pi_{\mathrm{im}(P)} Q \Pi_{\mathrm{im}(P)}) = \mathrm{F}(\Pi_{\mathrm{im}(Q)} P \Pi_{\mathrm{im}(Q)}, Q)$.
5. $\mathrm{F}(P, Q) \geq 0$, *with equality if and only if* $PQ = 0$.
6. $\mathrm{F}(P, Q)^2 \leq \mathrm{Tr}(P)\,\mathrm{Tr}(Q)$, *with equality if and only if* P *and* Q *are linearly dependent.*
7. *For every complex Euclidean space* $\mathcal{Y}$ *with* $\dim(\mathcal{Y}) \geq \dim(\mathcal{X})$ *and every isometry* $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$, *it holds that* $\mathrm{F}(P, Q) = \mathrm{F}(VPV^*, VQV^*)$.

Proof. The first three statements follow directly from the definition of the fidelity function: the fidelity function is defined as a composition of continuous functions, and is therefore continuous at every point in its domain; it holds that $\|A\|_1 = \|A^*\|_1$ for any choice of an operator A, and therefore

$$\left\| \sqrt{P}\sqrt{Q} \right\|_1 = \left\| \left(\sqrt{P}\sqrt{Q} \right)^* \right\|_1 = \left\| \sqrt{Q}\sqrt{P} \right\|_1; \tag{3.67}$$

and by the positive scalability of the trace norm, one has

$$\left\| \sqrt{\lambda P}\sqrt{Q} \right\|_1 = \sqrt{\lambda} \left\| \sqrt{P}\sqrt{Q} \right\|_1 = \left\| \sqrt{P}\sqrt{\lambda Q} \right\|_1. \tag{3.68}$$

Moving on to the fourth statement, it follows from the observation

$$\sqrt{P} = \sqrt{P}\,\Pi_{\mathrm{im}(P)} = \Pi_{\mathrm{im}(P)}\sqrt{P} \tag{3.69}$$

that

$$\sqrt{P}Q\sqrt{P} = \sqrt{P}\Pi_{\mathrm{im}(P)} Q \Pi_{\mathrm{im}(P)}\sqrt{P}. \tag{3.70}$$

Through the use of the expression (3.66), it follows that

$$\mathrm{F}(P, Q) = \mathrm{F}\big(P, \Pi_{\mathrm{im}(P)} Q \Pi_{\mathrm{im}(P)}\big). \tag{3.71}$$

This proves the first equality in statement 4, while the second equality follows through a combination of the first equality and statement 2.

Statement 5 follows from the fact that the trace norm is positive definite:

$$\left\| \sqrt{P}\sqrt{Q} \right\|_1 \geq 0, \tag{3.72}$$

with equality if and only if $\sqrt{P}\sqrt{Q} = 0$, which is equivalent to $PQ = 0$.

To prove the sixth statement, observe first that, by (1.176), there must exist a unitary operator $U \in \mathrm{U}(\mathcal{X})$ for which

$$\mathrm{F}(P,Q)^2 = \left\| \sqrt{P}\sqrt{Q} \right\|_1^2 = \left| \left\langle U, \sqrt{P}\sqrt{Q} \right\rangle \right|^2 = \left| \left\langle \sqrt{P}U, \sqrt{Q} \right\rangle \right|^2. \tag{3.73}$$

By the Cauchy–Schwarz inequality, it holds that

$$\left| \left\langle \sqrt{P}U, \sqrt{Q} \right\rangle \right|^2 \leq \left\| \sqrt{P}U \right\|_2^2 \left\| \sqrt{Q} \right\|_2^2 = \mathrm{Tr}(P)\,\mathrm{Tr}(Q), \tag{3.74}$$

which establishes the claimed inequality in statement 6. If it is the case that P and Q are linearly dependent, then it must hold that $P = \lambda Q$ or $Q = \lambda P$ for some choice of a nonnegative real number λ. In either case, it is straightforward to verify that

$$\mathrm{F}(P,Q)^2 = \mathrm{Tr}(P)\,\mathrm{Tr}(Q). \tag{3.75}$$

On the other hand, if P and Q are linearly independent, then so too are $\sqrt{P}U$ and $\sqrt{Q}$ for all unitary operators U; for if it holds that

$$\alpha\sqrt{P}U + \beta\sqrt{Q} = 0 \tag{3.76}$$

for scalars $\alpha, \beta \in \mathbb{C}$, then it follows that $|\alpha|^2 P = |\beta|^2 Q$. The assumption that P and Q are linearly independent therefore implies that a strict inequality occurs in the application of the Cauchy–Schwarz inequality in (3.74), which completes the proof of statement 6.

Finally, to prove statement 7, one may observe first that

$$\sqrt{VPV^*} = V\sqrt{P}V^* \quad \text{and} \quad \sqrt{VQV^*} = V\sqrt{Q}V^* \tag{3.77}$$

for every isometry $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$. By the isometric invariance of the trace norm, it follows that

$$\mathrm{F}(VPV^*, VQV^*) = \left\| V\sqrt{P}V^*V\sqrt{Q}V^* \right\|_1 = \left\| \sqrt{P}\sqrt{Q} \right\|_1, \tag{3.78}$$

which proves statement 7. □

Statements 5 and 6 of Proposition 3.12 imply that

$$0 \leq \mathrm{F}(\rho, \sigma) \leq 1 \tag{3.79}$$

for all density operators $\rho, \sigma \in \mathrm{D}(\mathcal{X})$. Moreover, $\mathrm{F}(\rho, \sigma) = 0$ if and only if ρ and σ have orthogonal images, and $\mathrm{F}(\rho, \sigma) = 1$ if and only if $\rho = \sigma$.

The output of the fidelity function is given by a simple formula when one of its input operators has rank equal to 1, as the next proposition states.

Proposition 3.13. *Let $\mathcal{X}$ be a complex Euclidean space, let $v \in \mathcal{X}$ be a vector, and let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator. It holds that*

$$\mathrm{F}\left(P, vv^*\right) = \sqrt{v^* P v}. \tag{3.80}$$

In particular, for every choice of vectors $u, v \in \mathcal{X}$, it holds that

$$\mathrm{F}\left(uu^*, vv^*\right) = |\langle u, v \rangle|. \tag{3.81}$$

Proof. The operator

$$\sqrt{P}\, vv^* \sqrt{P} \tag{3.82}$$

is positive semidefinite and has rank at most 1, which makes its vector of eigenvalues straightforward to calculate:

$$\lambda_1\Bigl(\sqrt{P}\, vv^* \sqrt{P}\Bigr) = \mathrm{Tr}\Bigl(\sqrt{P}\, vv^* \sqrt{P}\Bigr) = v^* P\, v \tag{3.83}$$

and

$$\lambda_k\Bigl(\sqrt{P}\, vv^* \sqrt{P}\Bigr) = 0 \tag{3.84}$$

for $k \geq 2$. It follows that

$$\mathrm{F}(P, vv^*) = \mathrm{Tr}\left(\sqrt{\sqrt{P}\, vv^* \sqrt{P}}\right) = \sqrt{\lambda_1\Bigl(\sqrt{P}\, vv^* \sqrt{P}\Bigr)} = \sqrt{v^* P\, v}, \tag{3.85}$$

as claimed. □

The following proposition is representative of another case in which the fidelity function has a simple formula. One corollary of this proposition, known as *Winter's gentle measurement lemma*, is useful in some situations.[2]

[2] The term *gentle measurement* reflects the observation that if a measurement of a particular state yields a particular outcome with very high probability, then a non-destructive analogue of that measurement causes only a small perturbation to the state in the event that the likely outcome is obtained.

Proposition 3.14. *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{F}(P, QPQ) = \langle P, Q \rangle. \tag{3.86}$$

Proof. It holds that

$$\sqrt{\sqrt{P}QPQ\sqrt{P}} = \sqrt{\left(\sqrt{P}Q\sqrt{P}\right)^2} = \sqrt{P}Q\sqrt{P}, \tag{3.87}$$

and therefore

$$\mathrm{F}(P, QPQ) = \mathrm{Tr}\left(\sqrt{\sqrt{P}QPQ\sqrt{P}}\right) = \mathrm{Tr}\left(\sqrt{P}Q\sqrt{P}\right) = \langle P, Q \rangle, \tag{3.88}$$

as claimed. □

Corollary 3.15 (Winter's gentle measurement lemma)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, and let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator satisfying $P \leq \mathbb{1}_{\mathcal{X}}$ and $\langle P, \rho \rangle > 0$. It holds that*

$$\mathrm{F}\left(\rho, \frac{\sqrt{P}\rho\sqrt{P}}{\langle P, \rho \rangle}\right) \geq \sqrt{\langle P, \rho \rangle}. \tag{3.89}$$

Proof. By Proposition 3.14, along with statement 3 of Proposition 3.12, one has

$$\mathrm{F}\left(\rho, \frac{\sqrt{P}\rho\sqrt{P}}{\langle P, \rho \rangle}\right) = \frac{1}{\sqrt{\langle P, \rho \rangle}} \mathrm{F}\left(\rho, \sqrt{P}\rho\sqrt{P}\right) = \frac{\langle \sqrt{P}, \rho \rangle}{\sqrt{\langle P, \rho \rangle}}. \tag{3.90}$$

Under the assumption $0 \leq P \leq \mathbb{1}$, it holds that $\sqrt{P} \geq P$, and therefore $\langle \sqrt{P}, \rho \rangle \geq \langle P, \rho \rangle$, from which the corollary follows. □

Another simple, yet very useful, property of the fidelity function is that it is multiplicative with respect to tensor products.

Proposition 3.16. *Let $P_0, Q_0 \in \mathrm{Pos}(\mathcal{X}_0)$ and $P_1, Q_1 \in \mathrm{Pos}(\mathcal{X}_1)$ be positive semidefinite operators, for complex Euclidean spaces $\mathcal{X}_0$ and $\mathcal{X}_1$. It holds that*

$$\mathrm{F}(P_0 \otimes P_1, Q_0 \otimes Q_1) = \mathrm{F}(P_0, Q_0)\,\mathrm{F}(P_1, Q_1). \tag{3.91}$$

Proof. Operator square roots and compositions respect tensor products, and the trace norm is multiplicative with respect to tensor products, so

$$\begin{aligned} \mathrm{F}(P_0 \otimes P_1, Q_0 \otimes Q_1) &= \left\| \sqrt{P_0 \otimes P_1} \sqrt{Q_0 \otimes Q_1} \right\|_1 \\ &= \left\| \sqrt{P_0}\sqrt{Q_0} \otimes \sqrt{P_1}\sqrt{Q_1} \right\|_1 = \left\| \sqrt{P_0}\sqrt{Q_0} \right\|_1 \left\| \sqrt{P_1}\sqrt{Q_1} \right\|_1 \\ &= \mathrm{F}(P_0, Q_0)\ \mathrm{F}(P_1, Q_1), \end{aligned} \tag{3.92}$$

as claimed. □

3.2.2 Alternative characterizations of the fidelity function

Multiple alternative characterizations of the fidelity function are known; a selection of such alternative characterizations is presented below. Some of these characterizations will allow for further properties of the fidelity function to be established, or will find other uses elsewhere in this book.

Block operator characterization

The first alternate characterization of the fidelity function to be presented is given by the following theorem. This characterization is particularly useful for establishing relevant properties of the fidelity function, including joint concavity in its arguments and monotonicity under the actions of channels, as will be described in the section following this one.

Theorem 3.17. *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{F}(P, Q) = \max\left\{ \left|\mathrm{Tr}(X)\right| \,:\, X \in \mathrm{L}(\mathcal{X}),\ \begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X}) \right\}. \tag{3.93}$$

The following lemma, which will find other uses elsewhere in this book, will be used to prove Theorem 3.17. The lemma is stated in slightly greater generality than is needed in the present context, in that it does not require P and Q to act on the same spaces, but there is no added difficulty in proving it with this greater generality.

Lemma 3.18. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $P \in \mathrm{Pos}(\mathcal{X})$ and $Q \in \mathrm{Pos}(\mathcal{Y})$ be positive semidefinite operators, and let $X \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be an operator. It holds that*

$$\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{Y}) \tag{3.94}$$

if and only if $X = \sqrt{P}K\sqrt{Q}$ for $K \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ satisfying $\|K\| \leq 1$.

Proof. Suppose first that $X = \sqrt{P}K\sqrt{Q}$ for $K \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ being an operator for which $\|K\| \leq 1$. It follows that $KK^* \leq \mathbb{1}_{\mathcal{X}}$, and therefore

$$0 \leq \begin{pmatrix} \sqrt{P}K \\ \sqrt{Q} \end{pmatrix} \begin{pmatrix} K^*\sqrt{P} & \sqrt{Q} \end{pmatrix} = \begin{pmatrix} \sqrt{P}KK^*\sqrt{P} & X \\ X^* & Q \end{pmatrix} \leq \begin{pmatrix} P & X \\ X^* & Q \end{pmatrix}. \tag{3.95}$$

For the reverse implication, assume

$$\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{Y}), \tag{3.96}$$

and define

$$K = \sqrt{P^+}X\sqrt{Q^+}. \tag{3.97}$$

It will be proved that $X = \sqrt{P}K\sqrt{Q}$ and $\|K\| \leq 1$. Observe first that, for every Hermitian operator $H \in \mathrm{Herm}(\mathcal{X})$, the block operator

$$\begin{pmatrix} H & 0 \\ 0 & \mathbb{1} \end{pmatrix} \begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \begin{pmatrix} H & 0 \\ 0 & \mathbb{1} \end{pmatrix} = \begin{pmatrix} HPH & HX \\ X^*H & Q \end{pmatrix} \tag{3.98}$$

is positive semidefinite. In particular, for $H = \Pi_{\ker(P)}$ being the projection onto the kernel of P, one has that the operator

$$\begin{pmatrix} 0 & \Pi_{\ker(P)}X \\ X^*\Pi_{\ker(P)} & Q \end{pmatrix} \tag{3.99}$$

is positive semidefinite, which implies that $\Pi_{\ker(P)}X = 0$, and therefore $\Pi_{\mathrm{im}(P)}X = X$. Through a similar argument, one finds that $X\Pi_{\mathrm{im}(Q)} = X$. It therefore follows that

$$\sqrt{P}K\sqrt{Q} = \Pi_{\mathrm{im}(P)}X\Pi_{\mathrm{im}(Q)} = X. \tag{3.100}$$

Next, note that

$$\begin{pmatrix} x^*Px & x^*Xy \\ y^*X^*x & y^*Qy \end{pmatrix} = \begin{pmatrix} x^* & 0 \\ 0 & y^* \end{pmatrix} \begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \geq 0 \tag{3.101}$$

for every choice of vectors $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Setting

$$x = \sqrt{P^{+}}u \quad \text{and} \quad y = \sqrt{Q^{+}}v \tag{3.102}$$

for arbitrarily chosen unit vectors $u \in \mathcal{X}$ and $v \in \mathcal{Y}$, one finds that

$$\begin{pmatrix} 1 & u^{*}Kv \\ v^{*}K^{*}u & 1 \end{pmatrix} \geq \begin{pmatrix} u^{*}\Pi_{\mathrm{im}(P)}u & u^{*}Kv \\ v^{*}K^{*}u & v^{*}\Pi_{\mathrm{im}(Q)}v \end{pmatrix} \geq 0 \tag{3.103}$$

and therefore $|u^{*}Kv| \leq 1$. As this inequality holds for all unit vectors u and v, it follows that $\|K\| \leq 1$, as required. □

Proof of Theorem 3.17. By Lemma 3.18, the expression on the right-hand side of the equation (3.93) may be written as

$$\max\left\{ \left| \mathrm{Tr}\left(\sqrt{P}K\sqrt{Q}\right) \right| \,:\, K \in \mathrm{L}(\mathcal{X}),\, \|K\| \leq 1 \right\}, \tag{3.104}$$

which is equivalent to

$$\max\left\{ \left| \left\langle K, \sqrt{P}\sqrt{Q} \right\rangle \right| \,:\, K \in \mathrm{L}(\mathcal{X}),\, \|K\| \leq 1 \right\}. \tag{3.105}$$

By the duality of the trace and spectral norms, as expressed by (1.167), one has

$$\left\| \sqrt{P}\sqrt{Q} \right\|_{1} = \mathrm{F}(P, Q), \tag{3.106}$$

which completes the proof. □

Remark 3.19. For any choice of operators $P, Q \in \mathrm{Pos}(\mathcal{X})$ and $X \in \mathrm{L}(\mathcal{X})$, and a scalar $\alpha \in \mathbb{C}$ satisfying $|\alpha| = 1$, it holds that

$$\begin{pmatrix} P & X \\ X^{*} & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X}) \tag{3.107}$$

if and only if

$$\begin{pmatrix} P & \alpha X \\ \overline{\alpha}X^{*} & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X}). \tag{3.108}$$

This fact follows from Lemma 3.18. Alternatively, one may conclude that (3.107) implies (3.108) through the equation

$$\begin{pmatrix} \mathbb{1} & 0 \\ 0 & \alpha\mathbb{1} \end{pmatrix}^{*} \begin{pmatrix} P & X \\ X^{*} & Q \end{pmatrix} \begin{pmatrix} \mathbb{1} & 0 \\ 0 & \alpha\mathbb{1} \end{pmatrix} = \begin{pmatrix} P & \alpha X \\ \overline{\alpha}X^{*} & Q \end{pmatrix}, \tag{3.109}$$

while the reverse implication is obtained similarly, through the equation

$$\begin{pmatrix} \mathbb{1} & 0 \\ 0 & \alpha\mathbb{1} \end{pmatrix} \begin{pmatrix} P & \alpha X \\ \overline{\alpha} X^* & Q \end{pmatrix} \begin{pmatrix} \mathbb{1} & 0 \\ 0 & \alpha\mathbb{1} \end{pmatrix}^* = \begin{pmatrix} P & X \\ X^* & Q \end{pmatrix}. \tag{3.110}$$

For any two positive semidefinite operators $P, Q \in \mathrm{Pos}(\mathcal{X})$, it therefore holds that the fidelity $\mathrm{F}(P, Q)$ is given by the expression

$$\max\left\{ \Re(\mathrm{Tr}(X)) \,:\, X \in \mathrm{L}(\mathcal{X}),\, \begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X}) \right\}. \tag{3.111}$$

Moreover, there must exist an operator $X \in \mathrm{L}(\mathcal{X})$ such that

$$\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X}) \tag{3.112}$$

and $\mathrm{F}(P, Q) = \mathrm{Tr}(X)$.

The characterization of the fidelity function established by Theorem 3.17 provides an expression of the fidelity $\mathrm{F}(P, Q)$ corresponding to the optimal value of a semidefinite program, as will now be explained. First, define a map $\Phi \in \mathrm{T}(\mathcal{X} \oplus \mathcal{X})$ as

$$\Phi \begin{pmatrix} X_0 & \cdot \\ \cdot & X_1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} X_0 & 0 \\ 0 & X_1 \end{pmatrix} \tag{3.113}$$

for every $X_0, X_1 \in \mathrm{L}(\mathcal{X})$, where the dots represent elements of $\mathrm{L}(\mathcal{X})$ that have no influence on the output of this map. One may verify that the map Φ is self-adjoint: $\Phi = \Phi^*$. Then, for a given choice of $P, Q \in \mathrm{Pos}(\mathcal{X})$, define Hermitian operators $A, B \in \mathrm{Herm}(\mathcal{X} \oplus \mathcal{X})$ as

$$A = \frac{1}{2} \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \quad \text{and} \quad B = \frac{1}{2} \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}. \tag{3.114}$$

The primal and dual optimization problems associated with the semidefinite program (Φ, A, B), after minor simplifications, are as follows:

Primal problem		Dual problem	
maximize:	$\frac{1}{2}\mathrm{Tr}(X) + \frac{1}{2}\mathrm{Tr}(X^*)$	minimize:	$\frac{1}{2}\langle P, Y_0 \rangle + \frac{1}{2}\langle Q, Y_1 \rangle$
subject to:	$\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \geq 0,$	subject to:	$\begin{pmatrix} Y_0 & -\mathbb{1} \\ -\mathbb{1} & Y_1 \end{pmatrix} \geq 0,$
	$X \in \mathrm{L}(\mathcal{X}).$		$Y_0, Y_1 \in \mathrm{Herm}(\mathcal{X}).$

The optimal primal value of this semidefinite program is equal to $\mathrm{F}(P, Q)$, as it is in agreement with the expression (3.111).

The primal problem is evidently feasible, as one may simply take $X = 0$ to obtain a primal feasible solution. The dual problem is strictly feasible: for any choice of $Y_0 > \mathbb{1}$ and $Y_1 > \mathbb{1}$, one has that the operator

$$\begin{pmatrix} Y_0 & -\mathbb{1} \\ -\mathbb{1} & Y_1 \end{pmatrix} \tag{3.115}$$

is positive definite. Strong duality therefore follows by Slater's theorem (Theorem 1.13).

Alberti's theorem

Given that the semidefinite program for the fidelity described above possesses the property of strong duality, its dual optimum must be equal to the primal optimum $\mathrm{F}(P, Q)$. The following theorem is a consequence of this observation.

Theorem 3.20. *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{F}(P, Q) = \inf\left\{ \frac{1}{2}\langle P, Y\rangle + \frac{1}{2}\langle Q, Y^{-1}\rangle \,:\, Y \in \mathrm{Pd}(\mathcal{X}) \right\}. \tag{3.116}$$

Proof. Through the use of Lemma 3.18, it is routine to verify that the block operator

$$\begin{pmatrix} Y_0 & -\mathbb{1} \\ -\mathbb{1} & Y_1 \end{pmatrix} \tag{3.117}$$

is positive semidefinite, for a given choice of $Y_0, Y_1 \in \mathrm{Herm}(\mathcal{X})$, if and only if both Y_0 and Y_1 are positive definite and satisfy $Y_1 \geq Y_0^{-1}$. Because Q is positive semidefinite, it holds that $\langle Q, Y_1\rangle \geq \langle Q, Y_0^{-1}\rangle$ provided $Y_0 > 0$ and $Y_1 \geq Y_0^{-1}$, so the dual problem associated to the semidefinite program (Φ, A, B) defined from P and Q as above is equivalent to a minimization of

$$\frac{1}{2}\langle P, Y\rangle + \frac{1}{2}\langle Q, Y^{-1}\rangle \tag{3.118}$$

over all positive definite operators $Y \in \mathrm{Pd}(\mathcal{X})$. As the optimal solution to this problem is equal to $\mathrm{F}(P, Q)$, the theorem follows. □

Theorem 3.20 implies the following corollary, which states a fact that is known as Alberti's theorem.[3]

Corollary 3.21 (Alberti's theorem). *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{F}(P,Q)^2 = \inf\Big\{ \langle P, Y\rangle \langle Q, Y^{-1}\rangle \,:\, Y \in \mathrm{Pd}(\mathcal{X}) \Big\}. \tag{3.119}$$

Proof. If either of P or Q is zero, the corollary is trivial, so it may be taken as an assumption that neither P nor Q is zero for the remainder of the proof.

The arithmetic-geometric mean inequality implies that

$$\frac{1}{2}\langle P, Y\rangle + \frac{1}{2}\langle Q, Y^{-1}\rangle \geq \sqrt{\langle P, Y\rangle \langle Q, Y^{-1}\rangle} \tag{3.120}$$

for every operator $Y \in \mathrm{Pd}(\mathcal{X})$. By Theorem 3.20, one concludes that

$$\inf\Big\{ \langle P, Y\rangle \langle Q, Y^{-1}\rangle \,:\, Y \in \mathrm{Pd}(\mathcal{X}) \Big\} \leq \mathrm{F}(P,Q)^2. \tag{3.121}$$

On the other hand, for any choice of $Y \in \mathrm{Pd}(\mathcal{X})$, it holds that

$$\sqrt{\langle P, Y\rangle \langle Q, Y^{-1}\rangle} = \sqrt{\langle P, \alpha Y\rangle \langle Q, (\alpha Y)^{-1}\rangle} \tag{3.122}$$

for every nonzero real number $\alpha \in \mathbb{R}$. In particular, for

$$\alpha = \sqrt{\frac{\langle Q, Y^{-1}\rangle}{\langle P, Y\rangle}}, \tag{3.123}$$

which has been selected so that $\langle P, \alpha Y\rangle = \langle Q, (\alpha Y)^{-1}\rangle$, one has

$$\begin{aligned} \sqrt{\langle P, Y\rangle \langle Q, Y^{-1}\rangle} &= \sqrt{\langle P, \alpha Y\rangle \langle Q, (\alpha Y)^{-1}\rangle} \\ &= \frac{1}{2}\langle P, \alpha Y\rangle + \frac{1}{2}\langle Q, (\alpha Y)^{-1}\rangle \geq \mathrm{F}(P,Q), \end{aligned} \tag{3.124}$$

and therefore

$$\inf\Big\{ \langle P, Y\rangle \langle Q, Y^{-1}\rangle \,:\, Y \in \mathrm{Pd}(\mathcal{X}) \Big\} \geq \mathrm{F}(P,Q)^2, \tag{3.125}$$

which completes the proof. □

[3] One may also prove that Corollary 3.21 implies Theorem 3.20, so the two facts are in fact equivalent.

It is possible to prove Theorem 3.21 directly, without making use of semidefinite programming duality, as the following proof demonstrates.

Alternative proof of Theorem 3.21. The special case in which $P = Q$ will be considered first. It this case, one aims to prove

$$\inf\left\{\frac{1}{2}\langle Y, P\rangle + \frac{1}{2}\langle Y^{-1}, P\rangle \,:\, Y \in \mathrm{Pd}(\mathcal{X})\right\} = \mathrm{Tr}(P). \tag{3.126}$$

As $Y = \mathbb{1}$ is positive definite, it is evident that the infimum in (3.126) is at most $\mathrm{Tr}(P)$, so it suffices to prove

$$\frac{1}{2}\langle Y, P\rangle + \frac{1}{2}\langle Y^{-1}, P\rangle \geq \mathrm{Tr}(P) \tag{3.127}$$

for every choice of $Y \in \mathrm{Pd}(\mathcal{X})$. As the operator

$$\frac{Y + Y^{-1}}{2} - \mathbb{1} = \frac{1}{2}\left(Y^{\frac{1}{2}} - Y^{-\frac{1}{2}}\right)^2 \tag{3.128}$$

is the square of a Hermitian operator, it must be positive semidefinite, and therefore

$$\frac{1}{2}\left\langle Y + Y^{-1}, P\right\rangle \geq \langle \mathbb{1}, P\rangle = \mathrm{Tr}(P) \tag{3.129}$$

for every positive semidefinite operator P. This proves that equation (3.126) holds, and therefore proves the theorem in the special case $P = Q$.

Toward the proof of the general case, suppose that P and Q are positive definite operators. Let

$$R = \sqrt{\sqrt{P}Q\sqrt{P}}, \tag{3.130}$$

and define a mapping $\Phi \in \mathrm{CP}(\mathcal{X})$ as

$$\Phi(X) = R^{-\frac{1}{2}}\sqrt{P}X\sqrt{P}R^{-\frac{1}{2}} \tag{3.131}$$

for every $X \in \mathrm{L}(\mathcal{X})$. For $Y \in \mathrm{Pd}(\mathcal{X})$ being any positive definite operator, it holds that

$$\left\langle \Phi(Y), R\right\rangle = \langle Y, P\rangle \quad \text{and} \quad \left\langle \Phi(Y)^{-1}, R\right\rangle = \left\langle Y^{-1}, Q\right\rangle, \tag{3.132}$$

and therefore

$$\inf_{Y \in \mathrm{Pd}(\mathcal{X})} \frac{\langle Y, P\rangle + \langle Y^{-1}, Q\rangle}{2} = \inf_{Y \in \mathrm{Pd}(\mathcal{X})} \frac{\langle \Phi(Y), R\rangle + \langle \Phi(Y)^{-1}, R\rangle}{2}. \tag{3.133}$$

Observing that, as Y ranges over all positive definite operators, so too does $\Phi(Y)$, one has that

$$\inf_{Y \in \mathrm{Pd}(\mathcal{X})} \frac{\langle Y, P \rangle + \langle Y^{-1}, Q \rangle}{2} = \mathrm{Tr}(R) = \mathrm{F}(P, Q) \tag{3.134}$$

by the special case considered in the initial part of the proof.

Finally, in the most general case in which $P, Q \in \mathrm{Pos}(\mathcal{X})$ may not be invertible, the theorem follows from a continuity argument. In greater detail, for every positive real number $\varepsilon > 0$, one has

$$\frac{1}{2}\big\langle Y, P \big\rangle + \frac{1}{2}\big\langle Y^{-1}, Q \big\rangle \leq \frac{1}{2}\big\langle Y, P + \varepsilon \mathbb{1} \big\rangle + \frac{1}{2}\big\langle Y^{-1}, Q + \varepsilon \mathbb{1} \big\rangle \tag{3.135}$$

for every choice of $Y \in \mathrm{Pd}(\mathcal{X})$. Taking the infimum over all positive definite operators $Y \in \mathrm{Pd}(\mathcal{X})$ yields the inequality

$$\inf_{Y \in \mathrm{Pd}(\mathcal{X})} \frac{\big\langle Y, P \big\rangle + \big\langle Y^{-1}, Q \big\rangle}{2} \leq \mathrm{F}(P + \varepsilon \mathbb{1}, Q + \varepsilon \mathbb{1}), \tag{3.136}$$

which holds by virtue of the fact that $P + \varepsilon \mathbb{1}$ and $Q + \varepsilon \mathbb{1}$ are necessarily positive definite. As this inequality holds for all $\varepsilon > 0$, it follows from the continuity of the fidelity function that

$$\inf_{Y \in \mathrm{Pd}(\mathcal{X})} \frac{\big\langle Y, P \big\rangle + \big\langle Y^{-1}, Q \big\rangle}{2} \leq \mathrm{F}(P, Q). \tag{3.137}$$

On the other hand, for each choice of $Y \in \mathrm{Pd}(\mathcal{X})$, one has

$$\frac{1}{2}\big\langle Y, P + \varepsilon \mathbb{1} \big\rangle + \frac{1}{2}\big\langle Y^{-1}, Q + \varepsilon \mathbb{1} \big\rangle \geq \mathrm{F}(P + \varepsilon \mathbb{1}, Q + \varepsilon \mathbb{1}) \tag{3.138}$$

for all $\varepsilon > 0$, and therefore the inequality

$$\frac{1}{2}\big\langle Y, P \big\rangle + \frac{1}{2}\big\langle Y^{-1}, Q \big\rangle \geq \mathrm{F}(P, Q) \tag{3.139}$$

follows from the continuity of the expressions on the two sides of this inequality. This is so for all $Y \in \mathrm{Pd}(\mathcal{X})$, and therefore

$$\inf_{Y \in \mathrm{Pd}(\mathcal{X})} \frac{\big\langle Y, P \big\rangle + \big\langle Y^{-1}, Q \big\rangle}{2} \geq \mathrm{F}(P, Q), \tag{3.140}$$

which completes the proof. □

Uhlmann's theorem

Uhlmann's theorem establishes a link between the fidelity function and the notion of a purification of a state (or of a positive semidefinite operator more generally), providing a characterization of the fidelity function that finds many uses in the theory of quantum information. The elementary lemma that follows will be used to prove this theorem.

Lemma 3.22. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $A, B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be operators. It holds that*

$$\mathrm{F}\big(AA^*, BB^*\big) = \big\|A^*B\big\|_1. \tag{3.141}$$

Proof. Using the polar decomposition, one may write

$$A = PU \quad \text{and} \quad B = QV, \tag{3.142}$$

for $P, Q \in \mathrm{Pos}(\mathcal{X})$ being positive semidefinite operators and $U, V \in \mathrm{U}(\mathcal{X})$ being unitary operators. Applying the unitary invariance of the trace norm to the definition of the fidelity function, one finds that

$$\mathrm{F}\big(AA^*, BB^*\big) = \mathrm{F}\big(P^2, Q^2\big) = \big\|PQ\big\|_1 = \big\|U^*PQV\big\|_1 = \big\|A^*B\big\|_1, \tag{3.143}$$

as required. □

Theorem 3.23 (Uhlmann's theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators having rank at most $\dim(\mathcal{Y})$, and let $u \in \mathcal{X} \otimes \mathcal{Y}$ satisfy $\mathrm{Tr}_{\mathcal{Y}}(uu^*) = P$. It holds that*

$$\mathrm{F}(P, Q) = \max\big\{|\langle u, v\rangle| \,:\, v \in \mathcal{X} \otimes \mathcal{Y},\, \mathrm{Tr}_{\mathcal{Y}}(vv^*) = Q\big\}. \tag{3.144}$$

Proof. Let $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be the operator for which $u = \mathrm{vec}(A)$, let $w \in \mathcal{X} \otimes \mathcal{Y}$ be a vector satisfying $Q = \mathrm{Tr}_{\mathcal{Y}}(ww^*)$, and let $B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be the operator for which $w = \mathrm{vec}(B)$. It follows by the unitary equivalence of purifications (Theorem 2.11) that

$$\begin{aligned}
&\max\big\{|\langle u, v\rangle| \,:\, v \in \mathcal{X} \otimes \mathcal{Y},\, \mathrm{Tr}_{\mathcal{Y}}(vv^*) = Q\big\} \\
&\qquad = \max\big\{|\langle u, (\mathbb{1}_{\mathcal{X}} \otimes U)w\rangle| \,:\, U \in \mathrm{U}(\mathcal{Y})\big\} \\
&\qquad = \max\big\{\big|\langle A, BU^{\mathsf{T}}\rangle\big| \,:\, U \in \mathrm{U}(\mathcal{Y})\big\} \\
&\qquad = \max\big\{\big|\big\langle \overline{U}, A^*B\big\rangle\big| \,:\, U \in \mathrm{U}(\mathcal{Y})\big\} \\
&\qquad = \big\|A^*B\big\|_1.
\end{aligned} \tag{3.145}$$

By Lemma 3.22, it holds that

$$\left\|A^*B\right\|_1 = \mathrm{F}(AA^*, BB^*) = \mathrm{F}(P, Q), \tag{3.146}$$

which completes the proof. □

It will be convenient later in the chapter to make use of the following corollary, which is essentially a rephrasing of Lemma 3.22.

Corollary 3.24. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $u, v \in \mathcal{X} \otimes \mathcal{Y}$ be vectors. It holds that*

$$\mathrm{F}\big(\mathrm{Tr}_{\mathcal{Y}}\big(uu^*\big), \mathrm{Tr}_{\mathcal{Y}}\big(vv^*\big)\big) = \left\|\mathrm{Tr}_{\mathcal{X}}\big(vu^*\big)\right\|_1. \tag{3.147}$$

Remark 3.25. Note that the partial traces on the left-hand side of (3.147) are taken over the space $\mathcal{Y}$, while the partial trace on the right-hand side is taken over $\mathcal{X}$.

Proof of Corollary 3.24. Let $A, B \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be the operators for which it holds that $u = \mathrm{vec}(A)$ and $v = \mathrm{vec}(B)$. By Lemma 3.22, one has

$$\begin{aligned} \mathrm{F}\big(\mathrm{Tr}_{\mathcal{Y}}\big(uu^*\big), \mathrm{Tr}_{\mathcal{Y}}\big(vv^*\big)\big) &= \mathrm{F}\big(AA^*, BB^*\big) \\ &= \left\|A^*B\right\|_1 = \left\|(A^*B)^{\mathsf{T}}\right\|_1 = \left\|\mathrm{Tr}_{\mathcal{X}}\big(vu^*\big)\right\|_1 \end{aligned} \tag{3.148}$$

as required. □

Bhattacharyya coefficient characterization

The last characterization of the fidelity function to be described in this section is based on a quantity known as the *Bhattacharyya coefficient*. For any alphabet Σ, and for vectors $u, v \in [0, \infty)^{\Sigma}$ having nonnegative real number entries, the Bhattacharyya coefficient $\mathrm{B}(u, v)$ of u and v is defined as

$$\mathrm{B}(u, v) = \sum_{a \in \Sigma} \sqrt{u(a)}\sqrt{v(a)}. \tag{3.149}$$

The Bhattacharyya coefficient relates to the fidelity between commuting operators in a straightforward way, as the following proposition describes.

Proposition 3.26. *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semdefinite operators satisfying $[P, Q] = 0$. It holds that*

$$\mathrm{F}(P, Q) = \mathrm{B}(\lambda(P), \lambda(Q)). \tag{3.150}$$

Proof. Let $n = \dim(\mathcal{X})$. Given that P and Q commute, and are positive semidefinite (and therefore normal) operators, Theorem 1.5 implies that there must exist an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$ for which

$$P = \sum_{k=1}^{n} \lambda_k(P)\, x_k x_k^* \quad \text{and} \quad Q = \sum_{k=1}^{n} \lambda_k(Q)\, x_k x_k^*. \tag{3.151}$$

It therefore holds that

$$\sqrt{P}\sqrt{Q} = \sum_{k=1}^{n} \sqrt{\lambda_k(P)}\sqrt{\lambda_k(Q)}\, x_k x_k^*, \tag{3.152}$$

so that

$$\mathrm{F}(P,Q) = \left\| \sqrt{P}\sqrt{Q} \right\|_1 = \sum_{k=1}^{n} \sqrt{\lambda_k(P)}\sqrt{\lambda_k(Q)} = \mathrm{B}(\lambda(P), \lambda(Q)), \tag{3.153}$$

as required. □

There exists a more interesting connection between the Bhattacharyya coefficient and the fidelity function, concerning the measurement statistics generated from arbitrary pairs of states. The following notation is helpful when explaining this connection: for a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ and positive semidefinite operators $P, Q \in \mathrm{Pos}(\mathcal{X})$, one defines

$$\mathrm{B}(P, Q \,|\, \mu) = \sum_{a \in \Sigma} \sqrt{\langle \mu(a), P \rangle}\sqrt{\langle \mu(a), Q \rangle}. \tag{3.154}$$

Equivalently,

$$\mathrm{B}(P, Q \,|\, \mu) = \mathrm{B}(u, v) \tag{3.155}$$

for $u, v \in [0, \infty)^{\Sigma}$ being the vectors defined as

$$u(a) = \langle \mu(a), P \rangle \quad \text{and} \quad v(a) = \langle \mu(a), Q \rangle \tag{3.156}$$

for each $a \in \Sigma$.

Theorem 3.27. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. For every choice of a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, it holds that*

$$\mathrm{F}(P,Q) \leq \mathrm{B}(P, Q \,|\, \mu). \tag{3.157}$$

Moreover, if it is the case that $|\Sigma| \geq \dim(\mathcal{X})$, then there exists a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ for which equality holds in (3.157).

Proof. Assume first that $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ is an arbitrary measurement, and let $U \in \mathrm{U}(\mathcal{X})$ be a unitary operator satisfying

$$\mathrm{F}(P,Q) = \left\| \sqrt{P}\sqrt{Q} \right\|_1 = \left\langle U, \sqrt{P}\sqrt{Q} \right\rangle. \tag{3.158}$$

By the triangle inequality followed by the Cauchy–Schwarz inequality, one finds that

$$\begin{aligned} \mathrm{F}(P,Q) &= \left\langle U, \sqrt{P}\sqrt{Q} \right\rangle = \sum_{a\in\Sigma} \left\langle U, \sqrt{P}\mu(a)\sqrt{Q} \right\rangle \\ &\leq \sum_{a\in\Sigma} \left| \left\langle \sqrt{\mu(a)}\sqrt{P}U, \sqrt{\mu(a)}\sqrt{Q} \right\rangle \right| \\ &\leq \sum_{a\in\Sigma} \sqrt{\langle \mu(a), P\rangle}\sqrt{\langle \mu(a), Q\rangle} = \mathrm{B}(P,Q\,|\,\mu). \end{aligned} \tag{3.159}$$

Next, it will be proved, under the assumption $|\Sigma| \geq \dim(\mathcal{X})$, that there exists a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ for which $\mathrm{F}(P,Q) = \mathrm{B}(P,Q\,|\,\mu)$. It suffices to prove that there is a measurement $\mu : \{1,\ldots,n\} \to \mathrm{Pos}(\mathcal{X})$ for which $\mathrm{F}(P,Q) = \mathrm{B}(P,Q\,|\,\mu)$, for $n = \dim(\mathcal{X})$.

Consider first the case in which P is invertible. Define

$$R = P^{-\frac{1}{2}} \left(\sqrt{P}Q\sqrt{P} \right)^{\frac{1}{2}} P^{-\frac{1}{2}}, \tag{3.160}$$

and let

$$R = \sum_{k=1}^{n} \lambda_k(R) u_k u_k^* \tag{3.161}$$

be a spectral decomposition of R. One may verify that $Q = RPR$, from which it follows that

$$\begin{aligned} &\sum_{k=1}^{n} \sqrt{\langle u_k u_k^*, P\rangle}\sqrt{\langle u_k u_k^*, Q\rangle} = \sum_{k=1}^{n} \sqrt{\langle u_k u_k^*, P\rangle}\sqrt{\langle u_k u_k^*, RPR\rangle} \\ &\quad = \sum_{k=1}^{n} \lambda_k(R)\langle u_k u_k^*, P\rangle = \langle R, P\rangle = \mathrm{Tr}\left(\sqrt{\sqrt{P}Q\sqrt{P}} \right) = \mathrm{F}(P,Q). \end{aligned} \tag{3.162}$$

The measurement $\mu : \{1,\ldots,n\} \to \mathrm{Pos}(\mathcal{X})$ defined by

$$\mu(k) = u_k u_k^* \tag{3.163}$$

for each $k \in \{1,\ldots,n\}$ therefore satisfies $\mathrm{F}(P,Q) = \mathrm{B}(P,Q\,|\,\mu)$.

Finally, the case in which $r = \operatorname{rank}(P) < n$ will be considered. Let $\Pi = \Pi_{\operatorname{im}(P)}$ denote the projection onto the image of P. By restricting one's attention to this subspace, the argument above may be seen to imply the existence of an orthonormal basis $\{u_1, \ldots, u_r\}$ for $\operatorname{im}(P)$ that satisfies

$$\mathrm{F}(P, \Pi Q \Pi) = \sum_{k=1}^{r} \sqrt{\langle u_k u_k^*, P \rangle} \sqrt{\langle u_k u_k^*, \Pi Q \Pi \rangle}. \tag{3.164}$$

Let $\{u_1, \ldots, u_n\}$ be any orthonormal basis of $\mathcal{X}$ that is obtained by completing the orthonormal set $\{u_1, \ldots, u_r\}$. As $\langle u_k u_k^*, P \rangle = 0$ for $k > r$ and $\langle u_k u_k^*, \Pi Q \Pi \rangle = \langle u_k u_k^*, Q \rangle$ for $k \leq r$, it follows that

$$\begin{aligned} &\sum_{k=1}^{n} \sqrt{\langle u_k u_k^*, P \rangle} \sqrt{\langle u_k u_k^*, Q \rangle} \\ &\quad = \sum_{k=1}^{r} \sqrt{\langle u_k u_k^*, P \rangle} \sqrt{\langle u_k u_k^*, \Pi Q \Pi \rangle} = \mathrm{F}(P, \Pi Q \Pi) = \mathrm{F}(P, Q), \end{aligned} \tag{3.165}$$

where the final equality holds by statement 4 of Proposition 3.12. Once again, the measurement $\mu : \{1, \ldots, n\} \to \operatorname{Pos}(\mathcal{X})$ defined by (3.163) for each $k \in \{1, \ldots, n\}$ satisfies $\mathrm{F}(P, Q) = \mathrm{B}(P, Q \,|\, \mu)$, which completes the proof. □

3.2.3 Further properties of the fidelity function

Various properties of the fidelity function can be established by means of the alternative characterizations presented in Section 3.2.2.

Joint concavity and monotonicity under the action of channels

The next theorem may be proved using the block operator characterization of the fidelity function (Theorem 3.17). As a corollary of this theorem, one finds that the fidelity function is *jointly concave* in its arguments.

Theorem 3.28. *Let $P_0, P_1, Q_0, Q_1 \in \operatorname{Pos}(\mathcal{X})$ be positive semidefinite operators, for $\mathcal{X}$ being a complex Euclidean space. It holds that*

$$\mathrm{F}(P_0 + P_1, Q_0 + Q_1) \geq \mathrm{F}(P_0, Q_0) + \mathrm{F}(P_1, Q_1). \tag{3.166}$$

Proof. By Theorem 3.17 (together with Remark 3.19), one may choose operators $X_0, X_1 \in \mathrm{L}(\mathcal{X})$ such that the block operators

$$\begin{pmatrix} P_0 & X_0 \\ X_0^* & Q_0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} P_1 & X_1 \\ X_1^* & Q_1 \end{pmatrix} \tag{3.167}$$

are both positive semidefinite, and such that

$$\mathrm{Tr}(X_0) = \mathrm{F}(P_0, Q_0) \quad \text{and} \quad \mathrm{Tr}(X_1) = \mathrm{F}(P_1, Q_1). \tag{3.168}$$

The sum of two positive semidefinite operators is positive semidefinite, and therefore

$$\begin{pmatrix} P_0 + P_1 & X_0 + X_1 \\ (X_0 + X_1)^* & Q_0 + Q_1 \end{pmatrix} = \begin{pmatrix} P_0 & X_0 \\ X_0^* & Q_0 \end{pmatrix} + \begin{pmatrix} P_1 & X_1 \\ X_1^* & Q_1 \end{pmatrix} \tag{3.169}$$

is positive semidefinite. Applying Theorem 3.17 again, one finds that

$$\mathrm{F}(P_0 + P_1, Q_0 + Q_1) \geq |\mathrm{Tr}(X_0 + X_1)| = \mathrm{F}(P_0, Q_0) + \mathrm{F}(P_1, Q_1), \tag{3.170}$$

as required. □

Corollary 3.29 (Joint concavity of the fidelity function). *Let $\mathcal{X}$ be a complex Euclidean space, let $\rho_0, \rho_1, \sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X})$ be density operators, and let $\lambda \in [0, 1]$. It holds that*

$$\begin{aligned} &\mathrm{F}(\lambda\rho_0 + (1-\lambda)\rho_1, \lambda\sigma_0 + (1-\lambda)\sigma_1) \\ &\qquad \geq \lambda\,\mathrm{F}(\rho_0, \sigma_0) + (1-\lambda)\,\mathrm{F}(\rho_1, \sigma_1). \end{aligned} \tag{3.171}$$

Proof. By Theorem 3.28, together with statement 3 of Proposition 3.12, it holds that

$$\begin{aligned} &\mathrm{F}(\lambda\rho_0 + (1-\lambda)\rho_1, \lambda\sigma_0 + (1-\lambda)\sigma_1) \\ &\qquad \geq \mathrm{F}(\lambda\rho_0, \lambda\sigma_0) + \mathrm{F}((1-\lambda)\rho_1, (1-\lambda)\sigma_1) \\ &\qquad = \lambda\,\mathrm{F}(\rho_0, \sigma_0) + (1-\lambda)\,\mathrm{F}(\rho_1, \sigma_1), \end{aligned} \tag{3.172}$$

as claimed. □

The joint concavity of the fidelity function implies that the fidelity function is concave in each of its arguments individually:

$$\mathrm{F}(\lambda\rho_0 + (1-\lambda)\rho_1, \sigma) \geq \lambda\,\mathrm{F}(\rho_0, \sigma) + (1-\lambda)\,\mathrm{F}(\rho_1, \sigma) \tag{3.173}$$

for all $\rho_0, \rho_1, \sigma \in \mathrm{D}(\mathcal{X})$ and $\lambda \in [0,1]$, and similar for concavity in the second argument rather than the first.

The *monotonicity* of the fidelity function under the action of channels is another fundamental property that may be established using the block operator characterization.

Theorem 3.30. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $P, Q \in \mathrm{Pos}(\mathcal{X})$, and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that*

$$\mathrm{F}(P, Q) \leq \mathrm{F}(\Phi(P), \Phi(Q)). \tag{3.174}$$

Proof. By Theorem 3.17, one may choose $X \in \mathrm{L}(\mathcal{X})$ so that

$$\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \tag{3.175}$$

is positive semidefinite and satisfies $|\mathrm{Tr}(X)| = \mathrm{F}(P, Q)$. By the complete positivity of Φ, the block operator

$$\begin{pmatrix} \Phi(P) & \Phi(X) \\ \Phi(X^*) & \Phi(Q) \end{pmatrix} = \begin{pmatrix} \Phi(P) & \Phi(X) \\ \Phi(X)^* & \Phi(Q) \end{pmatrix} \tag{3.176}$$

is positive semidefinite as well. Invoking Theorem 3.17 again, and using the fact that Φ is trace-preserving, it follows that

$$\mathrm{F}(\Phi(P), \Phi(Q)) \geq |\mathrm{Tr}(\Phi(X))| = |\mathrm{Tr}(X)| = \mathrm{F}(P, Q), \tag{3.177}$$

as required. □

Fidelity between extensions of operators

Suppose, for a given choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, that $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ are positive semidefinite operators and $Q_0 \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ is a positive semidefinite operator that extends P_0, meaning that $\mathrm{Tr}_{\mathcal{Y}}(Q_0) = P_0$. For every operator $Q_1 \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ satisfying $\mathrm{Tr}_{\mathcal{Y}}(Q_1) = P_1$, it follows from Theorem 3.30 that

$$\mathrm{F}(Q_0, Q_1) \leq \mathrm{F}(\mathrm{Tr}_{\mathcal{Y}}(Q_0), \mathrm{Tr}_{\mathcal{Y}}(Q_1)) = \mathrm{F}(P_0, P_1). \tag{3.178}$$

It is natural, in some situations, to consider the maximum value that the fidelity $\mathrm{F}(Q_0, Q_1)$ may take, over all choices of an operator $Q_1 \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ extending P_1. As the following theorem establishes, this maximum value is necessarily equal to $\mathrm{F}(P_0, P_1)$, irrespective of the choice of Q_0.

Theorem 3.31. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, and let $Q_0 \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ satisfy $\mathrm{Tr}_{\mathcal{Y}}(Q_0) = P_0$. It holds that*

$$\max\bigl\{\mathrm{F}(Q_0, Q_1) : Q_1 \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}),\ \mathrm{Tr}_{\mathcal{Y}}(Q_1) = P_1\bigr\} \;=\; \mathrm{F}(P_0, P_1). \tag{3.179}$$

Proof. Let $\mathcal{Z}$ be a complex Euclidean space with $\dim(\mathcal{Z}) = \dim(\mathcal{X} \otimes \mathcal{Y})$, and choose any vector $u_0 \in \mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}$ satisfying

$$\mathrm{Tr}_{\mathcal{Z}}(u_0 u_0^*) = Q_0. \tag{3.180}$$

As Q_0 is an extension of P_0, it follows that

$$\mathrm{Tr}_{\mathcal{Y}\otimes\mathcal{Z}}(u_0 u_0^*) = P_0. \tag{3.181}$$

By Uhlmann's theorem (Theorem 3.23), there exists a vector $u_1 \in \mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}$ so that

$$\mathrm{Tr}_{\mathcal{Y}\otimes\mathcal{Z}}(u_1 u_1^*) = P_1 \quad \text{and} \quad |\langle u_0, u_1\rangle| = \mathrm{F}(P_0, P_1). \tag{3.182}$$

By setting

$$Q_1 = \mathrm{Tr}_{\mathcal{Z}}(u_1 u_1^*) \tag{3.183}$$

and applying Theorem 3.30 (for the channel being the partial trace over $\mathcal{Z}$), one has

$$\begin{aligned} \mathrm{F}(Q_0, Q_1) &= \mathrm{F}(\mathrm{Tr}_{\mathcal{Z}}(u_0 u_0^*), \mathrm{Tr}_{\mathcal{Z}}(u_1 u_1^*)) \\ &\geq \mathrm{F}(u_0 u_0^*, u_1 u_1^*) = |\langle u_0, u_1\rangle| = \mathrm{F}(P_0, P_1). \end{aligned} \tag{3.184}$$

This demonstrates that the maximum in (3.179) is at least $\mathrm{F}(P_0, P_1)$. The maximum is at most $\mathrm{F}(P_0, P_1)$ by (3.178), and so the proof is complete. □

A sum-of-squares relationship for fidelity

The next theorem states a useful fact relating the fidelity between two fixed states and the sum of the squared-fidelities between these two states and a third.

Theorem 3.32. *Let $\mathcal{X}$ be a complex Euclidean space and let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be density operators. It holds that*

$$\max_{\sigma \in \mathrm{D}(\mathcal{X})} \Bigl(\mathrm{F}(\rho_0, \sigma)^2 + \mathrm{F}(\rho_1, \sigma)^2\Bigr) = 1 + \mathrm{F}(\rho_0, \rho_1). \tag{3.185}$$

Proof. The proof will make use of the fact that, for any two unit vectors u_0 and u_1, chosen from an arbitrary complex Euclidean space, there is a simple closed-form expression for the largest eigenvalue of the sum of the rank-one projections corresponding to these vectors:

$$\lambda_1\bigl(u_0u_0^* + u_1u_1^*\bigr) = 1 + \bigl|\langle u_0, u_1\rangle\bigr|. \tag{3.186}$$

There are two steps of the proof, both of which combine the expression (3.186) with Uhlmann's theorem (Theorem 3.23). The first step proves the existence of a density operator $\sigma \in \mathrm{D}(\mathcal{X})$ such that

$$\mathrm{F}(\rho_0, \sigma)^2 + \mathrm{F}(\rho_1, \sigma)^2 \geq 1 + \mathrm{F}(\rho_0, \rho_1). \tag{3.187}$$

Let $\mathcal{Y}$ be a space with $\dim(\mathcal{Y}) = \dim(\mathcal{X})$, and let $u_0, u_1 \in \mathcal{X} \otimes \mathcal{Y}$ be vectors satisfying the following equations:

$$\begin{aligned} \mathrm{Tr}_{\mathcal{Y}}(u_0u_0^*) &= \rho_0, \\ \mathrm{Tr}_{\mathcal{Y}}(u_1u_1^*) &= \rho_1, \\ \bigl|\langle u_0, u_1\rangle\bigr| &= \mathrm{F}(\rho_0, \rho_1). \end{aligned} \tag{3.188}$$

The fact that there exists such a choice of vectors follows from Uhlmann's theorem.

Let $v \in \mathcal{X} \otimes \mathcal{Y}$ be a unit eigenvector of the operator $u_0u_0^* + u_1u_1^*$ that corresponds to its largest eigenvalue, so that

$$v^*\bigl(u_0u_0^* + u_1u_1^*\bigr)v = 1 + \bigl|\langle u_0, u_1\rangle\bigr|, \tag{3.189}$$

and let

$$\sigma = \mathrm{Tr}_{\mathcal{Y}}(vv^*). \tag{3.190}$$

Using Uhlmann's theorem again, one has

$$\mathrm{F}(\rho_0, \sigma) \geq \bigl|\langle u_0, v\rangle\bigr| \quad \text{and} \quad \mathrm{F}(\rho_1, \sigma) \geq \bigl|\langle u_1, v\rangle\bigr|, \tag{3.191}$$

so that

$$\begin{aligned} \mathrm{F}(\rho_0, \sigma)^2 + \mathrm{F}(\rho_1, \sigma)^2 &\geq v^*\bigl(u_0u_0^* + u_1u_1^*\bigr)v \\ &= 1 + \bigl|\langle u_0, u_1\rangle\bigr| = 1 + \mathrm{F}(\rho_0, \rho_1), \end{aligned} \tag{3.192}$$

which proves the required inequality.

The second step of the proof is to establish that the inequality

$$\mathrm{F}(\rho_0, \sigma)^2 + \mathrm{F}(\rho_1, \sigma)^2 \leq 1 + \mathrm{F}(\rho_0, \rho_1) \tag{3.193}$$

holds for every $\sigma \in \mathrm{D}(\mathcal{X})$. Again, let $\mathcal{Y}$ be a complex Euclidean space with $\dim(\mathcal{Y}) = \dim(\mathcal{X})$, let $\sigma \in \mathrm{D}(\mathcal{X})$ be chosen arbitrarily, and choose $v \in \mathcal{X} \otimes \mathcal{Y}$ to be any unit vector satisfying

$$\mathrm{Tr}_{\mathcal{Y}}(vv^*) = \sigma. \tag{3.194}$$

Also let $u_0, u_1 \in \mathcal{X} \otimes \mathcal{Y}$ be unit vectors satisfying the following equations:

$$\begin{aligned} \mathrm{Tr}_{\mathcal{Y}}(u_0 u_0^*) &= \rho_0, \\ \mathrm{Tr}_{\mathcal{Y}}(u_1 u_1^*) &= \rho_1, \\ |\langle u_0, v\rangle| &= \mathrm{F}(\rho_0, \sigma), \\ |\langle u_1, v\rangle| &= \mathrm{F}(\rho_1, \sigma). \end{aligned} \tag{3.195}$$

As in the first step of the proof, the existence of such vectors is implied by Uhlmann's theorem. As v is a unit vector, it holds that

$$\begin{aligned} v^*(u_0 u_0^* + u_1 u_1^*)v &\leq \lambda_1(u_0 u_0^* + u_1 u_1^*) \\ &= 1 + |\langle u_0, u_1\rangle| \leq 1 + \mathrm{F}(\rho_0, \rho_1), \end{aligned} \tag{3.196}$$

where the last inequality is, once again, implied by Uhlmann's theorem. Therefore, one has

$$\mathrm{F}(\rho_0, \sigma)^2 + \mathrm{F}(\rho_1, \sigma)^2 = v^*(u_0 u_0^* + u_1 u_1^*)v \leq 1 + \mathrm{F}(\rho_0, \rho_1), \tag{3.197}$$

as required. □

Fidelity between inputs and outputs of completely positive maps

With respect to the storage and transmission of quantum information, the identity map represents an ideal quantum channel, as this channel causes no disturbance to the quantum states it acts upon. For this reason, it may be desirable to measure the similarity between a given channel of the form $\Phi \in \mathrm{C}(\mathcal{X})$ and the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{X})}$ in some settings.

One setting in which such a comparison is made arises in connection with quantum source coding (to be discussed in Section 5.3.2). Here, one is interested in the fidelity between the input and output states of a given channel $\Phi \in \mathrm{C}(\mathcal{X})$, under the assumption that the channel acts on a state $\sigma \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ that extends a known fixed state $\rho \in \mathrm{D}(\mathcal{X})$. The *mapping fidelity*, which is specified by the following definition, is representative of this situation when σ is taken as a purification of the state ρ.

Definition 3.33. Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{CP}(\mathcal{X})$ be a completely positive map, and let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator. The *mapping fidelity* of Φ with respect to P is defined as

$$\mathrm{F}(\Phi, P) = \mathrm{F}\big(uu^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(uu^*)\big) \tag{3.198}$$

for $u = \mathrm{vec}\big(\sqrt{P}\big)$.

The mapping fidelity is also called the *channel fidelity* when Φ is a channel and $P = \rho$ is a density operator. (It is also commonly called the *entanglement fidelity* in this case, although that terminology will not be used in this book.)

An explicit formula for the mapping fidelity $\mathrm{F}(\Phi, P)$, from any Kraus representation of the mapping Φ, is given by the following proposition.

Proposition 3.34. *Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{CP}(\mathcal{X})$ be a completely positive map, and assume that a Kraus representation of Φ is given:*

$$\Phi(X) = \sum_{a \in \Sigma} A_a X A_a^* \tag{3.199}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for $\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$ being a collection of operators indexed by some alphabet Σ. For every operator $P \in \mathrm{Pos}(\mathcal{X})$, it holds that

$$\mathrm{F}(\Phi, P) = \sqrt{\sum_{a \in \Sigma} \big|\langle P, A_a \rangle\big|^2}. \tag{3.200}$$

Proof. Using Proposition 3.13, one may evaluate the expression (3.198) to obtain

$$\begin{aligned} \mathrm{F}(\Phi, P) &= \sqrt{\sum_{a \in \Sigma} \big|\mathrm{vec}\big(\sqrt{P}\big)^* (A_a \otimes \mathbb{1}_{\mathcal{X}}) \,\mathrm{vec}\big(\sqrt{P}\big)\big|^2} \\ &= \sqrt{\sum_{a \in \Sigma} \big|\big\langle \sqrt{P}, A_a \sqrt{P} \big\rangle\big|^2} = \sqrt{\sum_{a \in \Sigma} \big|\langle P, A_a \rangle\big|^2}, \end{aligned} \tag{3.201}$$

as required. □

As the next proposition implies, the purification $u = \mathrm{vec}\big(\sqrt{P}\big)$ taken in the definition of the mapping fidelity is representative of a worst case scenario. That is, for an arbitrary state $\sigma \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ that extends a known fixed state $\rho \in \mathrm{D}(\mathcal{X})$, the fidelity

$$\mathrm{F}\big(\sigma, (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\sigma)\big) \tag{3.202}$$

can be no smaller than the mapping fidelity $\mathrm{F}(\Phi, \rho)$.

Proposition 3.35. *Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{CP}(\mathcal{X})$ be a completely positive map, and let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator. Suppose further that $\mathcal{Y}$ and $\mathcal{Z}$ are complex Euclidean spaces, $u \in \mathcal{X} \otimes \mathcal{Y}$ is a vector satisfying $\mathrm{Tr}_{\mathcal{Y}}(uu^*) = P$, and $Q \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Z})$ is an operator satisfying $\mathrm{Tr}_{\mathcal{Z}}(Q) = P$. It holds that*

$$\mathrm{F}(Q, (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(Q)) \geq \mathrm{F}(uu^*, (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(uu^*)). \tag{3.203}$$

Proof. By Proposition 2.29, there must exist a channel $\Psi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ such that

$$\big(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Psi\big)(uu^*) = Q. \tag{3.204}$$

By Theorem 3.30, one has

$$\begin{aligned} &\mathrm{F}\big(uu^*, (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(uu^*)\big) \\ &\quad \leq \mathrm{F}\big(\big(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Psi\big)(uu^*), (\Phi \otimes \Psi)(uu^*)\big) \\ &\quad = \mathrm{F}(Q, (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(Q)), \end{aligned} \tag{3.205}$$

which completes the proof. □

It is also evident from this proposition that taking any other purification of P in place of $u = \mathrm{vec}\big(\sqrt{P}\big)$ in Definition 3.33 would yield precisely the same value.

Fuchs–van de Graaf inequalities

The final property of the fidelity function to be established in this section concerns its connection to the trace distance between quantum states. This is an important relationship, as it allows for an approximate conversion between the more operationally motivated trace distance and the often more analytically robust fidelity function evaluated on a given pair of states.

Theorem 3.36 (Fuchs–van de Graaf inequalities)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $\rho, \sigma \in \mathrm{D}(\mathcal{X})$ be density operators. It holds that*

$$1 - \frac{1}{2}\big\|\rho - \sigma\big\|_1 \leq \mathrm{F}(\rho, \sigma) \leq \sqrt{1 - \frac{1}{4}\big\|\rho - \sigma\big\|_1^2}. \tag{3.206}$$

Equivalently,

$$2 - 2\,\mathrm{F}(\rho, \sigma) \leq \big\|\rho - \sigma\big\|_1 \leq 2\sqrt{1 - \mathrm{F}(\rho, \sigma)^2}. \tag{3.207}$$

Proof. The two inequalities in (3.207) will be established separately, beginning with the first. By Theorem 3.27, there exists an alphabet Σ and a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ such that

$$\mathrm{F}(\rho, \sigma) = \mathrm{B}(\rho, \sigma \,|\, \mu). \tag{3.208}$$

Fix such a measurement, and define probability vectors $p, q \in \mathcal{P}(\Sigma)$ as

$$p(a) = \langle \mu(a), \rho \rangle \quad \text{and} \quad q(a) = \langle \mu(a), \sigma \rangle \tag{3.209}$$

for each $a \in \Sigma$, so that $\mathrm{B}(p, q) = \mathrm{F}(\rho, \sigma)$. By Proposition 3.5, together with the observation that

$$\bigl(\sqrt{\alpha} - \sqrt{\beta}\bigr)^2 \leq |\alpha - \beta| \tag{3.210}$$

for every choice of nonnegative real numbers $\alpha, \beta \geq 0$, it follows that

$$\begin{aligned} \bigl\|\rho - \sigma\bigr\|_1 &\geq \bigl\|p - q\bigr\|_1 = \sum_{a \in \Sigma} |p(a) - q(a)| \\ &\geq \sum_{a \in \Sigma} \Bigl(\sqrt{p(a)} - \sqrt{q(a)}\Bigr)^2 = 2 - 2\,\mathrm{B}(p, q) = 2 - 2\,\mathrm{F}(\rho, \sigma). \end{aligned} \tag{3.211}$$

The first inequality in (3.207) is therefore proved.

Next, the second inequality in (3.207) will be proved. Let $\mathcal{Y}$ be a complex Euclidean space with $\dim(\mathcal{Y}) = \dim(\mathcal{X})$. It follows by Uhlmann's theorem (Theorem 3.23) that there exists a choice of unit vectors $u, v \in \mathcal{X} \otimes \mathcal{Y}$ satisfying the equations

$$\begin{aligned} \mathrm{Tr}_{\mathcal{Y}}(uu^*) &= \rho, \\ \mathrm{Tr}_{\mathcal{Y}}(vv^*) &= \sigma, \end{aligned} \tag{3.212}$$

and

$$|\langle u, v \rangle| = \mathrm{F}(\rho, \sigma). \tag{3.213}$$

By the identity (1.180), it holds that

$$\bigl\|uu^* - vv^*\bigr\|_1 = 2\sqrt{1 - |\langle u, v \rangle|^2} = 2\sqrt{1 - \mathrm{F}(\rho, \sigma)^2}. \tag{3.214}$$

Consequently, by the monotonicity of the trace norm under partial tracing (1.177), one has

$$\bigl\|\rho - \sigma\bigr\|_1 \leq \bigl\|uu^* - vv^*\bigr\|_1 = 2\sqrt{1 - \mathrm{F}(\rho, \sigma)^2}. \tag{3.215}$$

The second inequality in (3.207) has been established, which completes the proof. □

The use of the Bhattacharyya coefficient characterization of the fidelity (Theorem 3.27) in the above proof may be substituted by the following operator norm inequality, which is a useful inequality in its own right.

Lemma 3.37. *Let $\mathcal{X}$ be a complex Euclidean space and let $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\left\| P_0 - P_1 \right\|_1 \geq \left\| \sqrt{P_0} - \sqrt{P_1} \right\|_2^2. \tag{3.216}$$

Proof. Let

$$\sqrt{P_0} - \sqrt{P_1} = Q_0 - Q_1, \tag{3.217}$$

for $Q_0, Q_1 \in \mathrm{Pos}(\mathcal{X})$, be the Jordan-Hahn decomposition of $\sqrt{P_0} - \sqrt{P_1}$, and let $\Pi_0 = \mathrm{im}(Q_0)$ and $\Pi_1 = \mathrm{im}(Q_1)$. The operator $\Pi_0 - \Pi_1$ has spectral norm at most 1, and therefore

$$\left\| P_0 - P_1 \right\|_1 \geq \left\langle \Pi_0 - \Pi_1, P_0 - P_1 \right\rangle. \tag{3.218}$$

Through the use of the operator identity

$$A^2 - B^2 = \frac{1}{2}(A - B)(A + B) + \frac{1}{2}(A + B)(A - B), \tag{3.219}$$

one finds that

$$\begin{aligned}
&\left\langle \Pi_0 - \Pi_1, P_0 - P_1 \right\rangle \\
&\quad = \frac{1}{2}\Big\langle \Pi_0 - \Pi_1, \Big(\sqrt{P_0} - \sqrt{P_1}\Big)\Big(\sqrt{P_0} + \sqrt{P_1}\Big)\Big\rangle \\
&\qquad\quad + \frac{1}{2}\Big\langle \Pi_0 - \Pi_1, \Big(\sqrt{P_0} + \sqrt{P_1}\Big)\Big(\sqrt{P_0} - \sqrt{P_1}\Big)\Big\rangle \\
&\quad = \frac{1}{2}\,\mathrm{Tr}\Big((Q_0 + Q_1)\Big(\sqrt{P_0} + \sqrt{P_1}\Big)\Big) \\
&\qquad\quad + \frac{1}{2}\,\mathrm{Tr}\Big(\Big(\sqrt{P_0} + \sqrt{P_1}\Big)(Q_0 + Q_1)\Big) \\
&\quad = \Big\langle Q_0 + Q_1, \sqrt{P_0} + \sqrt{P_1}\Big\rangle.
\end{aligned} \tag{3.220}$$

Finally, as Q_0, Q_1, $\sqrt{P_0}$, and $\sqrt{P_1}$ are positive semidefinite, one has

$$\begin{aligned}
&\Big\langle Q_0 + Q_1, \sqrt{P_0} + \sqrt{P_1}\Big\rangle \\
&\qquad \geq \Big\langle Q_0 - Q_1, \sqrt{P_0} - \sqrt{P_1}\Big\rangle = \Big\| \sqrt{P_0} - \sqrt{P_1} \Big\|_2^2,
\end{aligned} \tag{3.221}$$

which completes the proof. □

Alternative proof of Theorem 3.36. For the first inequality in (3.207), one has

$$\begin{aligned} \bigl\|\rho - \sigma\bigr\|_1 &\geq \Bigl\|\sqrt{\rho} - \sqrt{\sigma}\Bigr\|_2^2 = \operatorname{Tr}\bigl(\sqrt{\rho} - \sqrt{\sigma}\bigr)^2 \\ &= 2 - 2\operatorname{Tr}\bigl(\sqrt{\rho}\sqrt{\sigma}\bigr) \geq 2 - 2\,\mathrm{F}(\rho,\sigma) \end{aligned} \tag{3.222}$$

by Lemma 3.37. The second inequality in (3.207) is proved as before. □

3.3 Channel distances and discrimination

The trace norm induces a notion of distance between quantum states that is closely related to the task of state discrimination, as established by the Holevo–Helstrom theorem (Theorem 3.4). The present section discusses an analogous notion of distance for channels, induced by a norm known as the *completely bounded trace norm*, along with a similar connection to the task of *channel discrimination*.

3.3.1 Channel discrimination

The task of discriminating between pairs of channels is represented by the scenario that follows.

Scenario 3.38. Let X and Y be registers, let Z be a classical register having classical state set $\{0,1\}$, let $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be channels, and let $\lambda \in [0,1]$ be a real number. The channels Φ_0 and Φ_1, as well as the number λ, are assumed to be known to both Alice and Bob.

Alice prepares Z in a probabilistic state, so that its state is 0 with probability λ and 1 with probability $1-\lambda$. Conditioned on the state of Z, Alice interacts with Bob in one of the following two ways:

1. If $\mathrm{Z} = 0$, Alice receives X from Bob, transforms X into Y according to the action of Φ_0, and sends Y to Bob.
2. If $\mathrm{Z} = 1$, Alice receives X from Bob, transforms X into Y according to the action of Φ_1, and sends Y to Bob.

Bob's goal is to determine the classical state of Z, by means of an interaction with Alice.

One approach Bob may choose to take in this scenario is to select a state $\sigma \in \mathrm{D}(\mathcal{X})$ that maximizes the quantity

$$\big\| \lambda\rho_0 - (1-\lambda)\rho_1 \big\|_1, \tag{3.223}$$

for $\rho_0 = \Phi_0(\sigma)$ and $\rho_1 = \Phi_1(\sigma)$. If he prepares the register X in the state σ and sends it to Alice, he will receive Y in either of the states ρ_0 or ρ_1, and can then measure Y using an optimal measurement for discriminating ρ_0 and ρ_1 given with probabilities λ and $1-\lambda$, respectively.

This, however, is not the most general approach. More generally, Bob may make use of an *auxiliary* register W in the following way. First, he prepares the pair of registers (X,W) in some chosen state $\sigma \in \mathrm{D}(\mathcal{X}\otimes\mathcal{W})$, and then he allows Alice to transform X into Y according to Φ_0 or Φ_1. This results in the pair (Y,W) being in one of the two states

$$\rho_0 = \big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\sigma) \quad \text{and} \quad \rho_1 = \big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\sigma), \tag{3.224}$$

with probabilities λ and $1-\lambda$, respectively. Finally, he measures the pair (Y,W) in order to discriminate these two states. This more general approach can, in some cases, result in a striking improvement in the probability to discriminate Φ_0 and Φ_1, as the following example illustrates.

Example 3.39. Let $n \geq 2$, let Σ be an alphabet with $|\Sigma| = n$, and let X be a register having classical state set Σ. Define two channels $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X})$ as follows:

$$\begin{aligned} \Phi_0(X) &= \frac{1}{n+1}\big((\operatorname{Tr} X)\mathbb{1} + X^{\mathsf{T}}\big), \\ \Phi_1(X) &= \frac{1}{n-1}\big((\operatorname{Tr} X)\mathbb{1} - X^{\mathsf{T}}\big), \end{aligned} \tag{3.225}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

The maps Φ_0 and Φ_1, which are sometimes called the *Werner–Holevo channels*, are indeed channels. These maps are evidently trace preserving, and the fact that they are completely positive follows from a calculation of their Choi representations:

$$J(\Phi_0) = \frac{\mathbb{1}\otimes\mathbb{1} + W}{n+1} \quad \text{and} \quad J(\Phi_1) = \frac{\mathbb{1}\otimes\mathbb{1} - W}{n-1}. \tag{3.226}$$

Here, $W \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$ is the swap operator, which satisfies $W(u \otimes v) = v \otimes u$ for every $u, v \in \mathcal{X}$. As W is unitary and Hermitian, the operators $J(\Phi_0)$ and $J(\Phi_1)$ are both positive semidefinite.

Now, consider the channels Φ_0 and Φ_1, along with the scalar value

$$\lambda = \frac{n+1}{2n}, \tag{3.227}$$

in Scenario 3.38. It holds that

$$\lambda\Phi_0(X) - (1-\lambda)\Phi_1(X) = \frac{1}{n}X^{\mathsf{T}} \tag{3.228}$$

for every $X \in \mathrm{L}(\mathcal{X})$, and therefore

$$\big\|\lambda\Phi_0(\sigma) - (1-\lambda)\Phi_1(\sigma)\big\|_1 = \frac{1}{n} \tag{3.229}$$

for every choice of a density operator $\sigma \in \mathrm{D}(\mathcal{X})$. This quantity is relatively small when n is large, which is consistent with the observation that $\Phi_0(\sigma)$ and $\Phi_1(\sigma)$ are both close to the completely mixed state for any choice of an input $\sigma \in \mathrm{D}(\mathcal{X})$. If Bob prepares X in some state σ, and elects not to use an auxiliary register W, his probability to correctly identify the classical state of Z is therefore at most

$$\frac{1}{2} + \frac{1}{2n}. \tag{3.230}$$

On the other hand, if Bob makes use of an auxiliary register, the situation is quite different. In particular, suppose that W is a register sharing the same classical state set Σ as X, and suppose that Bob prepares the pair (X, W) in the state $\tau \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W})$ defined as

$$\tau = \frac{1}{n}\sum_{a,b\in\Sigma} E_{a,b} \otimes E_{a,b}. \tag{3.231}$$

The actions of the channels Φ_0 and Φ_1 on this state are as follows:

$$\begin{aligned} \big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\tau) &= \frac{\mathbb{1}\otimes\mathbb{1} + W}{n^2+n}, \\ \big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\tau) &= \frac{\mathbb{1}\otimes\mathbb{1} - W}{n^2-n}. \end{aligned} \tag{3.232}$$

These are orthogonal density operators, following from the calculation

$$\langle \mathbb{1}\otimes\mathbb{1} + W, \mathbb{1}\otimes\mathbb{1} - W\rangle = \mathrm{Tr}\big(\mathbb{1}\otimes\mathbb{1} + W - W - W^2\big) = 0. \tag{3.233}$$

It is therefore the case that the states $\big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\tau)$ and $\big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\tau)$ can be discriminated without error: for every $\lambda \in [0,1]$, one has

$$\big\| \lambda\big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\tau) - (1-\lambda)\big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\tau) \big\|_1 = 1. \tag{3.234}$$

By making use of an auxiliary register W in this way, Bob can therefore correctly discriminate the channels Φ_0 and Φ_1 without error.

This example makes clear that auxiliary registers must be taken into account when considering the optimal probability with which channels can be discriminated.

3.3.2 The completely bounded trace norm

This section defines a norm on the space $\mathrm{T}(\mathcal{X}, \mathcal{Y})$, for $\mathcal{X}$ and $\mathcal{Y}$ being any two complex Euclidean spaces, known as the *completely bounded trace norm*, and establishes some of its properties. The precise connection between this norm and the task of channel discrimination will be explained in the section following this one, but it will be evident from its definition that this norm is motivated in part by the discussion from the previous section stressing the importance of auxiliary systems in the task of channel discrimination.

The induced trace norm

When introducing the completely bounded trace norm, it is appropriate to begin with the definition of a related norm known as the *induced trace norm*.

Definition 3.40. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces. For each map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, the *induced trace norm* of Φ is defined as

$$\big\|\Phi\big\|_1 = \max\Big\{ \big\|\Phi(X)\big\|_1 \,:\, X \in \mathrm{L}(\mathcal{X}),\ \ \big\|X\big\|_1 \leq 1 \Big\}. \tag{3.235}$$

True to its name, this norm is an example of an *induced norm*; in general, one may consider the norm obtained by replacing the two trace norms in this definition with any other choices of norms that are defined on $\mathrm{L}(\mathcal{X})$ and $\mathrm{L}(\mathcal{Y})$. The use of the maximum, rather than the supremum, is justified in this context by the observation that the norm defined on $\mathrm{L}(\mathcal{Y})$ is continuous and the unit ball with respect to the norm defined on $\mathrm{L}(\mathcal{X})$ is compact.

Generally speaking, the induced trace norm fails to provide a physically well-motivated measure of distance between channels. It will, nevertheless,

be useful to consider some basic properties of this norm, for many of these properties will be inherited by the completely bounded trace norm, to be defined shortly.

The first property of the induced trace norm to be observed is that the maximum in Definition 3.40 is always achieved by a rank-one operator X.

Proposition 3.41. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map. It holds that*

$$\big\|\Phi\big\|_1 = \max_{u,v \in \mathcal{S}(\mathcal{X})} \big\|\Phi(uv^*)\big\|_1. \tag{3.236}$$

Proof. Every operator in $X \in \mathrm{L}(\mathcal{X})$ satisfying $\|X\|_1 \leq 1$ can be written as a convex combination of operators of the form uv^*, for $u, v \in \mathcal{S}(\mathcal{X})$ being unit vectors. The equation (3.236) follows from the fact that the trace norm is a convex function. □

Under the additional assumption that the mapping under consideration is positive, one has that the maximum in Definition 3.40 is achieved by a rank-one projection, as the following theorem states.

Theorem 3.42 (Russo–Dye)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a positive map. It holds that*

$$\big\|\Phi\big\|_1 = \max_{u \in \mathcal{S}(\mathcal{X})} \mathrm{Tr}\big(\Phi(uu^*)\big). \tag{3.237}$$

Proof. Using the duality of the trace and spectral norms, along with the identity (1.176), one finds that

$$\big\|\Phi\big\|_1 = \max_{U \in \mathrm{U}(\mathcal{Y})} \big\|\Phi^*(U)\big\|. \tag{3.238}$$

Consider an arbitrary unitary operator $U \in \mathrm{U}(\mathcal{Y})$, and let

$$U = \sum_{k=1}^{m} \lambda_k \Pi_k \tag{3.239}$$

be the spectral decomposition of U. Define an operator $P_k = \Phi^*(\Pi_k)$ for each index $k \in \{1, \ldots, m\}$. As Φ is positive, it holds that Φ^* is also positive (by Proposition 2.17), and therefore $P_1, \ldots, P_m \in \mathrm{Pos}(\mathcal{X})$. By Lemma 3.3,

along with the obervation that the eigenvalues $\lambda_1, \ldots, \lambda_m$ all lie on the unit circle, it follows that

$$\big\|\Phi^*(U)\big\| = \left\|\sum_{k=1}^{m} \lambda_k P_k\right\| \leq \left\|\sum_{k=1}^{m} P_k\right\| = \big\|\Phi^*(\mathbb{1}_{\mathcal{Y}})\big\|. \tag{3.240}$$

Consequently, as $\mathbb{1}_{\mathcal{Y}}$ is itself a unitary operator, one has

$$\|\Phi\|_1 = \big\|\Phi^*(\mathbb{1}_{\mathcal{Y}})\big\|. \tag{3.241}$$

Finally, as $\Phi^*(\mathbb{1}_{\mathcal{Y}})$ is necessarily positive semidefinite, it follows that

$$\big\|\Phi^*(\mathbb{1}_{\mathcal{Y}})\big\| = \max_{u \in \mathcal{S}(\mathcal{X})} \big\langle uu^*, \Phi^*(\mathbb{1}_{\mathcal{Y}})\big\rangle = \max_{u \in \mathcal{S}(\mathcal{X})} \operatorname{Tr}\big(\Phi(uu^*)\big), \tag{3.242}$$

which completes the proof. □

Corollary 3.43. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a positive and trace-preserving map. It holds that $\big\|\Phi\big\|_1 = 1$.*

The next proposition establishes three basic properties of the induced trace norm: submultiplicativity under compositions, additivity of channel differences under compositions, and unitary invariance.

Proposition 3.44. *For every choice of complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, the following facts regarding the induced trace norm hold:*

1. *For all maps $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{T}(\mathcal{Y}, \mathcal{Z})$, it holds that*

$$\big\|\Psi\Phi\big\|_1 \leq \big\|\Psi\big\|_1 \big\|\Phi\big\|_1. \tag{3.243}$$

2. *For all channels $\Phi_0, \Psi_0 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Phi_1, \Psi_1 \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$, it holds that*

$$\big\|\Psi_1\Psi_0 - \Phi_1\Phi_0\big\|_1 \leq \big\|\Psi_0 - \Phi_0\big\|_1 + \big\|\Psi_1 - \Phi_1\big\|_1. \tag{3.244}$$

3. *Let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map, let $U_0, V_0 \in \mathrm{U}(\mathcal{X})$ and $U_1, V_1 \in \mathrm{U}(\mathcal{Y})$ be unitary operators, and let $\Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be defined as*

$$\Psi(X) = U_1\Phi(U_0 X V_0)V_1 \tag{3.245}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It holds that $\|\Psi\|_1 = \|\Phi\|_1$.

Proof. To prove the first fact, one may observe that $\|\Psi(Y)\|_1 \leq \|\Psi\|_1 \|Y\|_1$ for every $Y \in \mathrm{L}(\mathcal{Y})$, and therefore

$$\big\|\Psi(\Phi(X))\big\|_1 \leq \big\|\Psi\big\|_1 \big\|\Phi(X)\big\|_1 \tag{3.246}$$

for every $X \in \mathrm{L}(\mathcal{X})$. Taking the maximum over all $X \in \mathrm{L}(\mathcal{X})$ with $\|X\|_1 \leq 1$ yields the inequality (3.243).

To prove the second fact, one may apply the triangle inequality, the inequality (3.243), and Corollary 3.43, to obtain

$$\begin{aligned}
\big\|\Psi_1\Psi_0 - \Phi_1\Phi_0\big\|_1 &\leq \big\|\Psi_1\Psi_0 - \Psi_1\Phi_0\big\|_1 + \big\|\Psi_1\Phi_0 - \Phi_1\Phi_0\big\|_1 \\
&= \big\|\Psi_1(\Psi_0 - \Phi_0)\big\|_1 + \big\|(\Psi_1 - \Phi_1)\Phi_0\big\|_1 \\
&\leq \big\|\Psi_1\big\|_1 \big\|\Psi_0 - \Phi_0\big\|_1 + \big\|\Psi_1 - \Phi_1\big\|_1 \big\|\Phi_0\big\|_1 \\
&= \big\|\Psi_0 - \Phi_0\big\|_1 + \big\|\Psi_1 - \Phi_1\big\|_1 .
\end{aligned} \tag{3.247}$$

Finally, by the unitary invariance of the trace norm, it follows that

$$\begin{aligned}
\big\|\Psi(X)\big\|_1 = \big\|U_1\Phi(U_0XV_0)V_1\big\|_1 &= \big\|\Phi(U_0XV_0)\big\|_1 \\
&\leq \big\|\Phi\big\|_1 \big\|U_0XV_0\big\|_1 = \big\|\Phi\big\|_1 \big\|X\big\|_1
\end{aligned} \tag{3.248}$$

for all $X \in \mathrm{L}(\mathcal{X})$, and therefore $\|\Psi\|_1 \leq \|\Phi\|_1$. By observing that

$$\Phi(X) = U_1^*\Psi(U_0^*XV_0^*)V_1^* \tag{3.249}$$

for all $X \in \mathrm{L}(\mathcal{X})$, one finds that $\|\Phi\|_1 \leq \|\Psi\|_1$ through a similar argument, which proves the third fact. □

One undesirable property of the induced trace norm is that it fails to be multiplicative with respect to tensor products, as the following example (which is closely related to Example 3.39) illustrates.

Example 3.45. Let $n \geq 2$, let Σ be an alphabet with $|\Sigma| = n$, let $\mathcal{X} = \mathbb{C}^{\Sigma}$, and consider the transpose map $\mathrm{T} \in \mathrm{T}(\mathcal{X})$, defined as $\mathrm{T}(X) = X^{\mathsf{T}}$ for all $X \in \mathrm{L}(\mathcal{X})$. It is evident that $\|\mathrm{T}\|_1 = 1$, as $\|X\|_1 = \|X^{\mathsf{T}}\|_1$ for every operator $X \in \mathrm{L}(\mathcal{X})$, and it holds that $\|\mathbb{1}_{\mathrm{L}(\mathcal{X})}\|_1 = 1$. On the other hand, one has

$$\big\|\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big\|_1 = n. \tag{3.250}$$

To verify this claim, one may first consider the density operator

$$\tau = \frac{1}{n} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b} \in \mathrm{D}(\mathcal{X} \otimes \mathcal{X}), \tag{3.251}$$

which has trace norm equal to 1. It holds that

$$\big\| \big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(\tau) \big\|_1 = \frac{1}{n} \big\| W \big\|_1 = n \tag{3.252}$$

for $W \in \mathrm{U}(\mathcal{X} \otimes \mathcal{X})$ denoting the swap operator, and therefore

$$\big\| \mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \big\|_1 \geq n. \tag{3.253}$$

To prove that $\| \mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \|_1$ is no larger than n, one may first observe that the relationship (1.162) between the trace and Frobenius norms implies

$$\big\| \big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(X) \big\|_1 \leq n \big\| \big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(X) \big\|_2 \tag{3.254}$$

for every operator $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$. As the entries of the operators X and $(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(X)$ are equal, up to being shuffled by the transposition mapping, one has that

$$\big\| \big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(X) \big\|_2 = \big\| X \big\|_2. \tag{3.255}$$

Finally, by (1.163) it holds that $\| X \|_2 \leq \| X \|_1$, from which it follows that

$$\big\| \mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \big\|_1 \leq n. \tag{3.256}$$

Definition of the completely bounded trace norm

The *completely bounded trace norm* is defined below. In words, its value for a given map is simply the induced trace norm of that map tensored with the identity map on the same input space as the mapping itself.

Definition 3.46. For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, the *completely bounded trace norm* of a mapping $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is defined as

$$\big|\!\big|\!\big| \Phi \big|\!\big|\!\big|_1 = \big\| \Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \big\|_1. \tag{3.257}$$

As the discussion in Section 3.3.1 has suggested, this is the more relevant norm, when compared with the induced trace norm, within the context of the channel discrimination task. In essence, the completely bounded trace norm quantifies the effect that a map may have when it acts on just one tensor factor of a tensor product space (or, in more physical terms, just one part of a compound system), as opposed to the action of that map on its input space alone. As it turns out, this definition not only yields a norm that

is more relevant to the channel discrimination task, but also one possessing many interesting and desirable properties (including multiplicativity with respect to tensor products).

The specific choice to take the identity mapping on $\mathrm{L}(\mathcal{X})$, as opposed to $\mathrm{L}(\mathcal{Y})$, or $\mathrm{L}(\mathcal{Z})$ for some other complex Euclidean space $\mathcal{Z}$, is explained in greater detail below. In simple terms, the space $\mathcal{X}$ is sufficiently large, and just large enough in the worst case, that the value (3.257) does not change if the identity mapping on $\mathrm{L}(\mathcal{X})$ is replaced by the identity mapping on $\mathrm{L}(\mathcal{Z})$, for any complex Euclidean space $\mathcal{Z}$ having dimension at least as large as the dimension of $\mathcal{X}$.

Basic properties of the completely bounded trace norm

The proposition that follows, which is immediate from Propositions 3.41 and 3.44 and Corollary 3.43, summarizes some of the basic properties that the completely bounded trace norm inherits from the induced trace norm.

Proposition 3.47. *For every choice of complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, the following facts regarding the completely bounded trace norm hold:*

1. *For all maps $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, it holds that*

$$\bigl|\!\bigl|\!\bigl|\Phi\bigr|\!\bigr|\!\bigr|_1 = \max\Bigl\{ \bigl\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(uv^*) \bigr\|_1 \,:\, u, v \in \mathcal{S}(\mathcal{X} \otimes \mathcal{X}) \Bigr\}. \tag{3.258}$$

2. *For all channels $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, it holds that $\bigl|\!\bigl|\!\bigl|\Phi\bigr|\!\bigr|\!\bigr|_1 = 1$.*
3. *For all maps $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{T}(\mathcal{Y}, \mathcal{Z})$, it holds that*

$$\bigl|\!\bigl|\!\bigl|\Psi\Phi\bigr|\!\bigr|\!\bigr|_1 \leq \bigl|\!\bigl|\!\bigl|\Psi\bigr|\!\bigr|\!\bigr|_1 \bigl|\!\bigl|\!\bigl|\Phi\bigr|\!\bigr|\!\bigr|_1. \tag{3.259}$$

4. *For all channels $\Phi_0, \Psi_0 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Phi_1, \Psi_1 \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$, it holds that*

$$\bigl|\!\bigl|\!\bigl|\Psi_1\Psi_0 - \Phi_1\Phi_0\bigr|\!\bigr|\!\bigr|_1 \leq \bigl|\!\bigl|\!\bigl|\Psi_0 - \Phi_0\bigr|\!\bigr|\!\bigr|_1 + \bigl|\!\bigl|\!\bigl|\Psi_1 - \Phi_1\bigr|\!\bigr|\!\bigr|_1. \tag{3.260}$$

5. *Let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map, let $U_0, V_0 \in \mathrm{U}(\mathcal{X})$ and $U_1, V_1 \in \mathrm{U}(\mathcal{Y})$ be unitary operators, and let $\Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be defined as*

$$\Psi(X) = U_1 \Phi(U_0 X V_0) V_1 \tag{3.261}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It holds that $\bigl|\!\bigl|\!\bigl|\Psi\bigr|\!\bigr|\!\bigr|_1 = \bigl|\!\bigl|\!\bigl|\Phi\bigr|\!\bigr|\!\bigr|_1$.

The next lemma will be used multiple times to establish further properties of the completely bounded trace norm.

Lemma 3.48. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map. For every choice of unit vectors $x, y \in \mathcal{X} \otimes \mathcal{Z}$ there exist unit vectors $u, v \in \mathcal{X} \otimes \mathcal{X}$ such that*

$$\big\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(xy^*) \big\|_1 = \big\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(uv^*) \big\|_1. \tag{3.262}$$

If it is the case that $x = y$, then the equality (3.262) holds under the additional requirement that $u = v$.

Proof. In the case that $\dim(\mathcal{Z}) \leq \dim(\mathcal{X})$, the lemma is straightforward: for any choice of a linear isometry $U \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$, the vectors $u = (\mathbb{1}_{\mathcal{X}} \otimes U)x$ and $v = (\mathbb{1}_{\mathcal{X}} \otimes U)y$ satisfy the required conditions.

In the case that $\dim(\mathcal{Z}) > \dim(\mathcal{X})$, one may consider Schmidt decompositions

$$x = \sum_{k=1}^{n} \sqrt{p_k}\, x_k \otimes z_k \quad \text{and} \quad y = \sum_{k=1}^{n} \sqrt{q_k}\, y_k \otimes w_k \tag{3.263}$$

of x and y, for $n = \dim(\mathcal{X})$, from which a suitable choice for the vectors u and v is given by

$$u = \sum_{k=1}^{n} \sqrt{p_k}\, x_k \otimes x_k \quad \text{and} \quad v = \sum_{k=1}^{n} \sqrt{q_k}\, y_k \otimes y_k. \tag{3.264}$$

For linear isometries $U, V \in \mathrm{U}(\mathcal{X}, \mathcal{Z})$ defined as

$$U = \sum_{k=1}^{n} z_k x_k^* \quad \text{and} \quad V = \sum_{k=1}^{n} w_k y_k^*, \tag{3.265}$$

it holds that $x = (\mathbb{1}_{\mathcal{X}} \otimes U)u$ and $y = (\mathbb{1}_{\mathcal{X}} \otimes V)v$, and therefore

$$\begin{aligned} \big\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(xy^*) \big\|_1 &= \big\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})((\mathbb{1} \otimes U)uv^*(\mathbb{1} \otimes V^*)) \big\|_1 \\ &= \big\| (\mathbb{1} \otimes U)(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(uv^*)(\mathbb{1} \otimes V^*) \big\|_1 \\ &= \big\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(uv^*) \big\|_1, \end{aligned} \tag{3.266}$$

as required. In case $x = y$, one may take the same Schmidt decomposition for x and y in (3.263), implying that $u = v$. □

With Lemma 3.48 in hand, the following theorem may be proved. The theorem implies a claim that was made earlier: the identity map on $\mathrm{L}(\mathcal{X})$ in Definition 3.46 could be replaced by the identity map on $\mathrm{L}(\mathcal{Z})$, for any space $\mathcal{Z}$ having dimension at least that of $\mathcal{X}$, without changing the value of the norm.

Theorem 3.49. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map, and let $\mathcal{Z}$ be a complex Euclidean space. It holds that*

$$\big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1 \leq \big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1, \tag{3.267}$$

with equality holding under the assumption that $\dim(\mathcal{Z}) \geq \dim(\mathcal{X})$.

Proof. By Proposition 3.41, there exist unit vectors $x, y \in \mathcal{X} \otimes \mathcal{Z}$ such that

$$\big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1 = \big\|\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(xy^*)\big\|_1. \tag{3.268}$$

Therefore, by Lemma 3.48, there exist unit vectors $u, v \in \mathcal{X} \otimes \mathcal{X}$ such that

$$\big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1 = \big\|\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(uv^*)\big\|_1, \tag{3.269}$$

which implies

$$\big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1 \leq \big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1. \tag{3.270}$$

To prove that equality holds under the assumption $\dim(\mathcal{Z}) \geq \dim(\mathcal{X})$, one may take $V \in \mathrm{U}(\mathcal{X}, \mathcal{Z})$ to be any isometry, and observe that

$$\begin{aligned}
\big\|\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(X)\big\|_1 &= \big\|(\mathbb{1}_{\mathcal{Y}} \otimes V)\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(X)(\mathbb{1}_{\mathcal{Y}} \otimes V)^*\big\|_1 \\
&= \big\|\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)\big((\mathbb{1}_{\mathcal{X}} \otimes V)X(\mathbb{1}_{\mathcal{X}} \otimes V)^*\big)\big\|_1 \\
&\leq \big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1 \big\|(\mathbb{1}_{\mathcal{X}} \otimes V)X(\mathbb{1}_{\mathcal{X}} \otimes V)^*\big\|_1 \\
&= \big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1 \big\|X\big\|_1 \\
&\leq \big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1,
\end{aligned} \tag{3.271}$$

for every operator $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$ with $\|X\|_1 \leq 1$, by the isometric invariance of the trace norm. It therefore holds that

$$\big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 \leq \big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1 \tag{3.272}$$

as required. □

Corollary 3.50. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map. It holds that*

$$\bigl|\!\bigl|\!\bigl|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\bigr|\!\bigr|\!\bigr|_1 = \bigl|\!\bigl|\!\bigl|\Phi\bigr|\!\bigr|\!\bigr|_1. \tag{3.273}$$

The fact that the completely bounded trace norm is multiplicative with respect to tensor products may now be proved.

Theorem 3.51. *Let $\Phi_0 \in \mathrm{T}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{T}(\mathcal{X}_1, \mathcal{Y}_1)$ be maps, for $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ being complex Euclidean spaces. It holds that*

$$\bigl|\!\bigl|\!\bigl|\Phi_0 \otimes \Phi_1\bigr|\!\bigr|\!\bigr|_1 = \bigl|\!\bigl|\!\bigl|\Phi_0\bigr|\!\bigr|\!\bigr|_1 \bigl|\!\bigl|\!\bigl|\Phi_1\bigr|\!\bigr|\!\bigr|_1. \tag{3.274}$$

Proof. By Proposition 3.47 and Corollary 3.50, it follows that

$$\begin{aligned} \bigl|\!\bigl|\!\bigl|\Phi_0 \otimes \Phi_1\bigr|\!\bigr|\!\bigr|_1 &= \bigl|\!\bigl|\!\bigl|(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y}_1)})(\mathbb{1}_{\mathrm{L}(\mathcal{X}_0)} \otimes \Phi_1)\bigr|\!\bigr|\!\bigr|_1 \\ &\leq \bigl|\!\bigl|\!\bigl|\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y}_1)}\bigr|\!\bigr|\!\bigr|_1 \bigl|\!\bigl|\!\bigl|\mathbb{1}_{\mathrm{L}(\mathcal{X}_0)} \otimes \Phi_1\bigr|\!\bigr|\!\bigr|_1 = \bigl|\!\bigl|\!\bigl|\Phi_0\bigr|\!\bigr|\!\bigr|_1 \bigl|\!\bigl|\!\bigl|\Phi_1\bigr|\!\bigr|\!\bigr|_1. \end{aligned} \tag{3.275}$$

It remains to prove the reverse inequality.

First, choose an operator $X_0 \in \mathrm{L}(\mathcal{X}_0 \otimes \mathcal{X}_0)$ such that $\|X_0\|_1 = 1$ and

$$\bigl|\!\bigl|\!\bigl|\Phi_0\bigr|\!\bigr|\!\bigr|_1 = \bigl\|\bigl(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_0)}\bigr)(X_0)\bigr\|_1, \tag{3.276}$$

as well as an operator $X_1 \in \mathrm{L}(\mathcal{X}_1 \otimes \mathcal{X}_1)$ such that $\|X_1\|_1 = 1$ and

$$\bigl|\!\bigl|\!\bigl|\Phi_1\bigr|\!\bigr|\!\bigr|_1 = \bigl\|\bigl(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_1)}\bigr)(X_1)\bigr\|_1. \tag{3.277}$$

The trace norm is multiplicative with respect to tensor products, and therefore $\|X_0 \otimes X_1\|_1 = 1$.

Next, observe that

$$\begin{aligned} \bigl|\!\bigl|\!\bigl|\Phi_0 \otimes \Phi_1\bigr|\!\bigr|\!\bigr|_1 &= \bigl\|\Phi_0 \otimes \Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_0 \otimes \mathcal{X}_1)}\bigr\|_1 \\ &= \bigl\|\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_0)} \otimes \Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_1)}\bigr\|_1. \end{aligned} \tag{3.278}$$

The second equality follows from the unitary invariance of the induced trace norm (the third statement of Proposition 3.44), which implies that this norm is invariant under permuting the ordering of tensor factors of maps. Again using the multiplicativity of the trace norm with respect to tensor products, it follows that

$$\begin{aligned} \bigl|\!\bigl|\!\bigl|\Phi_0 \otimes \Phi_1\bigr|\!\bigr|\!\bigr|_1 &\geq \bigl\|\bigl(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_0)} \otimes \Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_1)}\bigr)(X_0 \otimes X_1)\bigr\|_1 \\ &= \bigl\|\bigl(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_0)}\bigr)(X_0)\bigr\|_1 \bigl\|\bigl(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X}_1)}\bigr)(X_1)\bigr\|_1 \\ &= \bigl|\!\bigl|\!\bigl|\Phi_0\bigr|\!\bigr|\!\bigr|_1 \bigl|\!\bigl|\!\bigl|\Phi_1\bigr|\!\bigr|\!\bigr|_1, \end{aligned} \tag{3.279}$$

which completes the proof. □

3.3.3 Distances between channels

This section explains the connection between the completely bounded trace norm and the task of channel discrimination that was alluded to above, and discusses other aspects of the notion of distance between channels induced by the completely bounded trace norm.

The completely bounded trace norm of Hermiticity-preserving maps

For a given map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, one has that

$$\bigl|\!\bigl|\!\bigl|\Phi\bigr|\!\bigr|\!\bigr|_1 = \bigl\|(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(uv^*)\bigr\|_1 \tag{3.280}$$

for some choice of unit vectors $u, v \in \mathcal{X} \otimes \mathcal{X}$. The stronger condition that

$$\bigl|\!\bigl|\!\bigl|\Phi\bigr|\!\bigr|\!\bigr|_1 = \bigl\|(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(uu^*)\bigr\|_1 \tag{3.281}$$

for a single unit vector $u \in \mathcal{X} \otimes \mathcal{X}$ does not generally hold; without any restrictions on Φ, this could not reasonably be expected.

When the map Φ is Hermiticity-preserving, however, there will always exist a unit vector $u \in \mathcal{X} \otimes \mathcal{X}$ for which (3.281) holds. This fact is stated as Theorem 3.53 below, whose proof makes use of the following lemma.

Lemma 3.52. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a Hermiticity-preserving map, and let $\mathcal{Z}$ be any complex Euclidean space with $\dim(\mathcal{Z}) \geq 2$. There exists a unit vector $u \in \mathcal{X} \otimes \mathcal{Z}$ such that*

$$\bigl\|\bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\bigr)(uu^*)\bigr\|_1 \geq \bigl\|\Phi\bigr\|_1. \tag{3.282}$$

Proof. Let $X \in \mathrm{L}(\mathcal{X})$ be an operator for which it holds that $\|X\|_1 = 1$ and $\|\Phi(X)\|_1 = \|\Phi\|_1$. Let $z_0, z_1 \in \mathcal{Z}$ be any to orthogonal unit vectors, define a Hermitian operator $H \in \mathrm{Herm}(\mathcal{X} \otimes \mathcal{Z})$ as

$$H = \frac{1}{2} X \otimes z_0 z_1^* + \frac{1}{2} X^* \otimes z_1 z_0^*, \tag{3.283}$$

and observe that $\|H\|_1 = \|X\|_1 = 1$. Moreover, one has

$$\begin{aligned} \bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\bigr)(H) &= \frac{1}{2}\Phi(X) \otimes z_0 z_1^* + \frac{1}{2}\Phi(X^*) \otimes z_1 z_0^* \\ &= \frac{1}{2}\Phi(X) \otimes z_0 z_1^* + \frac{1}{2}\Phi(X)^* \otimes z_1 z_0^*, \end{aligned} \tag{3.284}$$

where the second equality follows from Theorem 2.25, together with the assumption that Φ is a Hermiticity-preserving map. It is therefore the case that

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(H) \big\|_1 = \big\| \Phi(X) \big\|_1 = \big\| \Phi \big\|_1. \tag{3.285}$$

Now consider a spectral decomposition

$$H = \sum_{k=1}^{n} \lambda_k u_k u_k^* \tag{3.286}$$

for $n = \dim(\mathcal{X} \otimes \mathcal{Z})$. By the triangle inequality, one has

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(H) \big\|_1 \leq \sum_{k=1}^{n} |\lambda_k| \, \big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(u_k u_k^*) \big\|_1. \tag{3.287}$$

As $\|H\|_1 = 1$, the expression on the right-hand side of the inequality (3.287) is a convex combination of the values

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(u_k u_k^*) \big\|_1, \tag{3.288}$$

ranging over $k \in \{1, \ldots, n\}$. There must therefore exist $k \in \{1, \ldots, n\}$ for which the inequality

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(u_k u_k^*) \big\|_1 \geq \big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(H) \big\|_1 = \big\| \Phi \big\|_1 \tag{3.289}$$

is satisfied. Setting $u = u_k$ completes the proof. □

Theorem 3.53. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a Hermiticity-preserving map. It holds that*

$$\big\| \big\| \Phi \big\| \big\|_1 = \max_{u \in \mathcal{S}(\mathcal{X} \otimes \mathcal{X})} \big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(uu^*) \big\|_1. \tag{3.290}$$

Proof. For every unit vector $u \in \mathcal{X} \otimes \mathcal{X}$, it holds that

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(uu^*) \big\|_1 \leq \big\| \Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \big\|_1 = \big\| \big\| \Phi \big\| \big\|_1, \tag{3.291}$$

so it suffices to prove that there exists a unit vector $u \in \mathcal{X} \otimes \mathcal{X}$ for which

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(uu^*) \big\|_1 \geq \big\| \Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \big\|_1 = \big\| \big\| \Phi \big\| \big\|_1. \tag{3.292}$$

Let $\mathcal{Z} = \mathbb{C}^2$. By Lemma 3.52 there exists a unit vector $x \in \mathcal{X} \otimes \mathcal{X} \otimes \mathcal{Z}$ such that

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(xx^*) \big\|_1 \geq \big\| \Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \big\|_1, \tag{3.293}$$

and by Lemma 3.48 there must exists a unit vector $u \in \mathcal{X} \otimes \mathcal{X}$ such that

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(uu^*) \big\|_1 = \big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(xx^*) \big\|_1. \tag{3.294}$$

For such a choice of u, one has (3.292), which completes the proof. □

A channel analogue of the Holevo–Helstrom theorem

The next theorem represents an analogue of the Holevo–Helstrom theorem (Theorem 3.4) for channels rather than states, with the completely bounded trace norm replacing the trace norm accordingly.

Theorem 3.54. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be channels, and let $\lambda \in [0,1]$. For any choice of a complex Euclidean space $\mathcal{Z}$, a density operator $\sigma \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$, and a measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$, it holds that*

$$\begin{aligned} \lambda \langle \mu(0), (\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) \rangle + (1-\lambda) \langle \mu(1), (\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) \rangle \\ \leq \frac{1}{2} + \frac{1}{2} \bigl| \bigl| \bigl| \lambda \Phi_0 - (1-\lambda) \Phi_1 \bigr| \bigr| \bigr|_1. \end{aligned} \tag{3.295}$$

Moreover, for any choice of $\mathcal{Z}$ satisfying $\dim(\mathcal{Z}) \geq \dim(\mathcal{X})$, equality is achieved in (3.295) for some choice of a pure state σ and a projective measurement μ.

Proof. By the Holevo–Helstrom theorem (Theorem 3.4), the quantity on the left-hand side of (3.295) is at most

$$\frac{1}{2} + \frac{1}{2} \bigl\| \lambda (\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) - (1-\lambda)(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) \bigr\|_1. \tag{3.296}$$

This value is upper-bounded by

$$\frac{1}{2} + \frac{1}{2} \bigl\| (\lambda \Phi_0 - (1-\lambda)\Phi_1) \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})} \bigr\|_1, \tag{3.297}$$

which is at most

$$\frac{1}{2} + \frac{1}{2} \bigl| \bigl| \bigl| \lambda \Phi_0 - (1-\lambda) \Phi_1 \bigr| \bigr| \bigr|_1 \tag{3.298}$$

by Theorem 3.49.

The mapping $\lambda \Phi_0 - (1-\lambda)\Phi_1$ is Hermiticity-preserving, by virtue of the fact that Φ_0 and Φ_1 are completely positive and λ is a real number. By Theorem 3.53, there must therefore exist a unit vector $u \in \mathcal{X} \otimes \mathcal{X}$ for which

$$\begin{aligned} \bigl\| \lambda (\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(uu^*) - (1-\lambda)(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(uu^*) \bigr\|_1 \\ = \bigl| \bigl| \bigl| \lambda \Phi_0 - (1-\lambda) \Phi_1 \bigr| \bigr| \bigr|_1. \end{aligned} \tag{3.299}$$

Under the assumption that $\dim(\mathcal{Z}) \geq \dim(\mathcal{X})$, one therefore has

$$\begin{aligned} \bigl\| \lambda (\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) - (1-\lambda)(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) \bigr\|_1 \\ = \bigl| \bigl| \bigl| \lambda \Phi_0 - (1-\lambda) \Phi_1 \bigr| \bigr| \bigr|_1 \end{aligned} \tag{3.300}$$

for

$$\sigma = (\mathbb{1}_{\mathcal{X}} \otimes V) u u^* (\mathbb{1}_{\mathcal{X}} \otimes V^*), \tag{3.301}$$

for an arbitrary choice of an isometry $V \in \mathrm{U}(\mathcal{X}, \mathcal{Z})$.

Finally, by the Holevo–Helstrom theorem, there must exist a projective measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$ such that

$$\begin{aligned} &\lambda \langle \mu(0), (\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) \rangle + (1-\lambda) \langle \mu(1), (\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) \rangle \\ &\qquad = \frac{1}{2} + \frac{1}{2} \big\| \lambda (\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\sigma) - (1-\lambda)(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(\sigma) \big\|_1 \\ &\qquad = \frac{1}{2} + \frac{1}{2} \big\| \lambda \Phi_0 - (1-\lambda) \Phi_1 \big\|_1, \end{aligned} \tag{3.302}$$

which completes the proof. □

Distances between networks of channels

Many computations and interactions that arise in the study of quantum information and computation can be represented as *networks* of channels. Here, one supposes that a collection of channels $\Phi_1, \ldots, \Phi_N$ having varying input and output spaces are arranged in an acyclic network, as suggested by the example depicted in Figure 3.1. The completely bounded trace norm is well-suited to analyses concerning errors, inaccuracies, and noise that may occur in such networks.

By composing the channels $\Phi_1, \ldots, \Phi_N$ in a manner consistent with the network, a single channel Φ is obtained. Under the assumption that registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are treated as inputs to the network and registers $\mathsf{Y}_1, \ldots, \mathsf{Y}_m$ are output, the channel Φ representing the composition of the channels $\Phi_1, \ldots, \Phi_N$ takes the form

$$\Phi \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m). \tag{3.303}$$

Now suppose that $\Psi_1, \ldots, \Psi_N$ are channels whose input spaces and output spaces agree with $\Phi_1, \ldots, \Phi_N$, respectively, and that Ψ_k is substituted for Φ_k for all $k = 1, \ldots, N$. Equivalently, the channels $\Psi_1, \ldots, \Psi_N$ are composed in a manner that is consistent with the description of the network, yielding a single channel

$$\Psi \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) \tag{3.304}$$

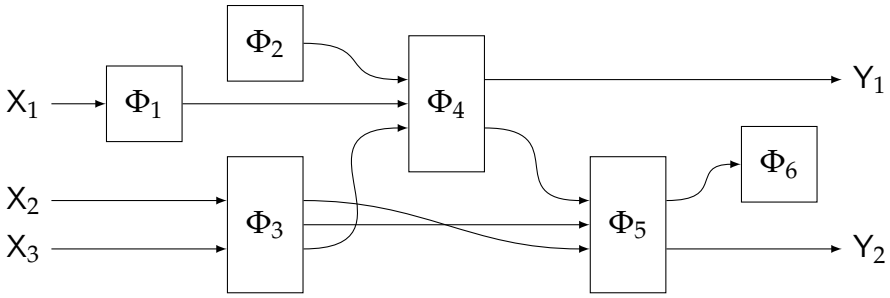


Figure 3.1: A hypothetical example of an acyclic network of channels. The arrows represent registers, and one assumes that the input and output spaces of the channels (represented by rectangles in the figure) are compatible with the registers represented by the arrows. For instance, the channel Φ_1 transforms the register X_1 into some other register (not explicitly named in the figure), which is the second of three inputs to the channel Φ_4. By composing the channels $\Phi_1, \ldots, \Phi_6$, one obtains a single channel $\Phi \in \mathrm{C}(\mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathcal{X}_3, \mathcal{Y}_1 \otimes \mathcal{Y}_2)$.

in place of Φ. It is natural to ask how much Φ and Ψ may differ, as a function of the differences between Φ_k and Ψ_k for $k = 1, \ldots, N$. It could be, for instance, that $\Phi_1, \ldots, \Phi_N$ represent ideal channels that are specified by a protocol or algorithm while $\Psi_1, \ldots, \Psi_N$ represent slightly noisy or corrupted variants of $\Phi_1, \ldots, \Phi_N$.

The following upper bound on the difference between Φ and Ψ, as a function of the differences between Φ_k and Ψ_k (for $k = 1, \ldots, N$), is obtained by induction from statement 4 of Proposition 3.47 along with Corollary 3.50:

$$\bigl|\!\bigl|\!\bigl|\Phi - \Psi\bigr|\!\bigr|\!\bigr|_1 \leq \sum_{k=1}^{N} \bigl|\!\bigl|\!\bigl|\Phi_k - \Psi_k\bigr|\!\bigr|\!\bigr|_1. \tag{3.305}$$

Thus, irrespective of the specific properties of the network under consideration, the differences between the channels Φ_k and Ψ_k for $k = 1, \ldots, N$ can only accumulate additively when they are composed into a network.

Discrimination between pairs of isometric channels

As Example 3.39 illustrates, it is necessary in some instances of Scenario 3.38 for Bob to use an auxiliary register W in order to optimally discriminate a given pair of channels. One interesting case in which it is *not* necessary for

Bob to make use of an auxiliary register in this scenario is when the two channels are isometric channels, defined as

$$\Phi_0(X) = V_0 X V_0^* \quad \text{and} \quad \Phi_1(X) = V_1 X V_1^* \tag{3.306}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for some choice of isometries $V_0, V_1 \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$. The fact that an auxiliary register is not needed for an optimal discrimination in this case is proved below. The proof makes use of the notion of the *numerical range* of an operator.

Definition 3.55. Let $\mathcal{X}$ be a complex Euclidean space and let $A \in \mathrm{L}(\mathcal{X})$ be an operator. The *numerical range* of A is the set $\mathcal{N}(A) \subset \mathbb{C}$ defined as follows:

$$\mathcal{N}(A) = \bigl\{ u^* A u \,:\, u \in \mathcal{S}(\mathcal{X}) \bigr\}. \tag{3.307}$$

In general, every eigenvalue of a given operator A is contained in $\mathcal{N}(A)$, and one may prove that $\mathcal{N}(A)$ is equal to the convex hull of the eigenvalues of A in the case that A is normal. For non-normal operators, however, this will not generally be the case. It is, however, always the case that $\mathcal{N}(A)$ is compact and convex, which is the content of the following theorem.

Theorem 3.56 (Toeplitz–Hausdorff theorem)**.** *For any complex Euclidean space $\mathcal{X}$ and any operator $A \in \mathrm{L}(\mathcal{X})$, the set $\mathcal{N}(A)$ is compact and convex.*

Proof. The function $f : \mathcal{S}(\mathcal{X}) \to \mathbb{C}$ defined by $f(u) = u^* A u$ is continuous, and the unit sphere $\mathcal{S}(\mathcal{X})$ is compact. Continuous functions map compact sets to compact sets, implying that $\mathcal{N}(A) = f(\mathcal{S}(\mathcal{X}))$ is compact.

It remains to prove that $\mathcal{N}(A)$ is convex. Fix any choice of $\alpha, \beta \in \mathcal{N}(A)$ and a real number $\lambda \in [0, 1]$. It will be proved that

$$\lambda\alpha + (1 - \lambda)\beta \in \mathcal{N}(A), \tag{3.308}$$

which suffices to prove the theorem. It will be assumed hereafter that $\alpha \neq \beta$, as the assertion is trivial in the case that $\alpha = \beta$.

By the definition of the numerical range, one may choose unit vectors $u, v \in \mathcal{S}(\mathcal{X})$ such that $u^* A u = \alpha$ and $v^* A v = \beta$. It follows from the fact that $\alpha \neq \beta$ that the vectors u and v are linearly independent.

Next, define

$$B = \frac{-\beta}{\alpha - \beta} \mathbb{1}_{\mathcal{X}} + \frac{1}{\alpha - \beta} A \tag{3.309}$$

so that $u^*Bu = 1$ and $v^*Bv = 0$. Let

$$X = \frac{B + B^*}{2} \quad \text{and} \quad Y = \frac{B - B^*}{2i} \tag{3.310}$$

represent the Hermitian and anti-Hermitian parts of B. It follows that

$$\begin{aligned} u^*Xu &= 1, \qquad v^*Xv = 0, \\ u^*Yu &= 0, \qquad v^*Yv = 0. \end{aligned} \tag{3.311}$$

Without loss of generality, it may be assumed that u^*Yv is purely imaginary (i.e., has real part equal to 0), for otherwise v may be replaced by $e^{i\theta}v$ for an appropriate choice of θ without changing any of the previously observed properties.

As u and v are linearly independent, the vector $tu + (1-t)v$ is nonzero for every choice of $t \in \mathbb{R}$. Thus, for each $t \in [0,1]$, one may define a unit vector

$$z(t) = \frac{tu + (1-t)v}{\|tu + (1-t)v\|}. \tag{3.312}$$

Because $u^*Yu = v^*Yv = 0$ and u^*Yv is purely imaginary, it follows that $z(t)^*Yz(t) = 0$ for every $t \in [0,1]$, and therefore

$$z(t)^*Bz(t) = z(t)^*Xz(t) = \frac{t^2 + 2t(1-t)\Re(v^*Xu)}{\|tu + (1-t)v\|}. \tag{3.313}$$

The expression on the right-hand side of (3.313) is a continuous real-valued function mapping 0 to 0 and 1 to 1. Consequently, there must exist at least one choice of $t \in [0,1]$ such that $z(t)^*Bz(t) = \lambda$. Let $w = z(t)$ for such choice of t, so that $w^*Bw = \lambda$. It holds that w is a unit vector, and

$$\begin{aligned} w^*Aw &= (\alpha - \beta)\left(\frac{\beta}{\alpha - \beta} + w^*Bw\right) \\ &= \beta + \lambda(\alpha - \beta) = \lambda\alpha + (1-\lambda)\beta. \end{aligned} \tag{3.314}$$

It has therefore been shown that $\lambda\alpha + (1-\lambda)\beta \in \mathcal{N}(A)$ as required. □

Theorem 3.57. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $V_0, V_1 \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$ be isometries, and let $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be channels defined as*

$$\Phi_0(X) = V_0XV_0^* \quad \textit{and} \quad \Phi_1(X) = V_1XV_1^* \tag{3.315}$$

for all $X \in \mathrm{L}(\mathcal{X})$. *There exists a unit vector* $u \in \mathcal{X}$ *such that*

$$\big\| \lambda \Phi_0(uu^*) - (1-\lambda)\Phi_1(uu^*) \big\|_1 = \big\| \lambda \Phi_0 - (1-\lambda)\Phi_1 \big\|_1 \tag{3.316}$$

for every $\lambda \in [0,1]$.

Proof. Using the identity (1.178), one finds that

$$\begin{aligned} &\big\| \lambda \Phi_0(uu^*) - (1-\lambda)\Phi_1(uu^*) \big\|_1 \\ &\qquad = \sqrt{1 - 4\lambda(1-\lambda)\left|u^* V_0^* V_1 u\right|^2}, \end{aligned} \tag{3.317}$$

for every unit vector $u \in \mathcal{X}$, and similarly

$$\begin{aligned} &\big\| \lambda \big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(vv^*) - (1-\lambda)\big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(vv^*) \big\|_1 \\ &\qquad = \sqrt{1 - 4\lambda(1-\lambda)\left|v^* \big(V_0^* V_1 \otimes \mathbb{1}_{\mathcal{Z}}\big) v\right|^2} \end{aligned} \tag{3.318}$$

for every complex Euclidean space $\mathcal{Z}$ and unit vector $v \in \mathcal{X} \otimes \mathcal{Z}$. Taking $\mathcal{Z}$ be a complex Euclidean space with $\dim(\mathcal{Z}) = \dim(\mathcal{X})$, it follows from (3.318) (along with Theorem 3.53) that there exists a unit vector $v \in \mathcal{X} \otimes \mathcal{Z}$ such that

$$\big\| \lambda \Phi_0 - (1-\lambda)\Phi_1 \big\|_1 = \sqrt{1 - 4\lambda(1-\lambda)\left|v^* \big(V_0^* V_1 \otimes \mathbb{1}_{\mathcal{Z}}\big) v\right|^2}. \tag{3.319}$$

Now, one may observe that

$$v^* \big(V_0^* V_1 \otimes \mathbb{1}_{\mathcal{Z}}\big) v = \big\langle \rho, V_0^* V_1 \big\rangle \tag{3.320}$$

for $\rho = \mathrm{Tr}_{\mathcal{Z}}(vv^*)$. By considering a spectral decomposition of the density operator ρ, one finds that the value represented by (3.320) is a convex combination of values of the form

$$w^* V_0^* V_1 w, \tag{3.321}$$

in which $w \in \mathcal{X}$ ranges over a set of unit eigenvectors of ρ. Each of these values is contained in the numerical range of $V_0^* V_1$, and therefore, by the Toeplitz–Hausdorff theorem (Theorem 3.56), there must exist a unit vector $u \in \mathcal{X}$ such that

$$u^* V_0^* V_1 u = \big\langle \rho, V_0^* V_1 \big\rangle. \tag{3.322}$$

By (3.317), it follows that

$$\big\| \lambda \Phi_0(uu^*) - (1-\lambda)\Phi_1(uu^*) \big\|_1 = \big\| \lambda \Phi_0 - (1-\lambda)\Phi_1 \big\|_1. \tag{3.323}$$

Observing that the vector u does not depend on λ, the proof is complete. □

The completely bounded trace distance from a channel to the identity

Returning once again to Example 3.39, one has that the Werner–Holevo channels can be perfectly discriminated through the use of a sufficiently large auiliary register, but are nearly indistinguishable (when the channels themselves are defined with respect to sufficiently large registers) without the use of an auxiliary register. The Werner–Holevo channels happen to have another feature, which is that they are highly noisy channels; their outputs are close to the completely mixed state for every possible input state.

One may ask if a similar phenomenon, in which an auxiliary register has a dramatic effect on the optimal probability of successfully discriminating channels, occurs when one of the channels is the identity channel. This is a natural question, as the closeness of a given channel to the identity channel may be a highly relevant figure of merit of that channel in some situations. The following theorem demonstrates that the phenomenon suggested above is limited in this setting. In particular, the theorem demonstrates that the potential advantage of using an auxiliary register in discriminating a given channel from the identity channel is dimension-independent.

Theorem 3.58. *Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{C}(\mathcal{X})$ be a channel, let $\varepsilon \in [0,1]$, and suppose that*

$$\big\| \Phi(\rho) - \rho \big\|_1 \leq \varepsilon \tag{3.324}$$

for every density operator $\rho \in \mathrm{D}(\mathcal{X})$. It holds that

$$\big\| \Phi - \mathbb{1}_{\mathrm{L}(\mathcal{X})} \big\|_1 \leq 2\sqrt{\varepsilon}. \tag{3.325}$$

Proof. It is evident from the assumptions of the theorem that, for every unit vector $u \in \mathcal{X}$, one has

$$\big\| \Phi(uu^*) - uu^* \big\|_1 \leq \varepsilon, \tag{3.326}$$

and therefore

$$\big| \big\langle uu^*, \Phi(uu^*) - uu^* \big\rangle \big| \leq \varepsilon. \tag{3.327}$$

The first main step of the proof will be to establish a bound of a similar nature:

$$\big| \big\langle uv^*, \Phi(uv^*) - uv^* \big\rangle \big| \leq \varepsilon, \tag{3.328}$$

for every pair of orthogonal unit vectors $u, v \in \mathcal{X}$. Toward this goal, assume that $u, v \in \mathcal{X}$ are orthogonal unit vectors, and define a unit vector

$$w_k = \frac{u + i^k v}{\sqrt{2}} \tag{3.329}$$

for each $k \in \{0,1,2,3\}$. From the observation that

$$uv^* = \frac{1}{2}\sum_{k=0}^{3} i^k w_k w_k^*, \tag{3.330}$$

it follows that

$$\Phi(uv^*) - uv^* = \frac{1}{2}\sum_{k=0}^{3} i^k \big(\Phi(w_k w_k^*) - w_k w_k^*\big). \tag{3.331}$$

Because the spectral norm of a traceless Hermitian operator is at most one-half of its trace norm, it follows that

$$\begin{aligned} \big\|\Phi(uv^*) - uv^*\big\| &\leq \frac{1}{2}\sum_{k=0}^{3} \big\|\Phi(w_k w_k^*) - w_k w_k^*\big\| \\ &\leq \frac{1}{4}\sum_{k=0}^{3} \big\|\Phi(w_k w_k^*) - w_k w_k^*\big\|_1 \leq \varepsilon. \end{aligned} \tag{3.332}$$

This implies the desired bound (3.328).

Now, let $z \in \mathcal{X} \otimes \mathcal{X}$ be a unit vector, expressed in the form of a Schmidt decomposition

$$z = \sum_{a\in\Sigma} \sqrt{p(a)}\, x_a \otimes y_a, \tag{3.333}$$

for Σ being an alphabet, $\{x_a : a \in \Sigma\}$ and $\{y_a : a \in \Sigma\}$ being orthonormal subsets of $\mathcal{X}$, and $p \in \mathcal{P}(\Sigma)$ being a probability vector. It holds that

$$\big\langle zz^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(zz^*)\big\rangle = \sum_{a,b\in\Sigma} p(a)p(b)\big\langle x_a x_b^*, \Phi(x_a x_b^*)\big\rangle, \tag{3.334}$$

and therefore, by the triangle inequality and the bounds (3.327) and (3.328) from above,

$$\begin{aligned} 1 - \big\langle zz^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(zz^*)\big\rangle &= \big|\big\langle zz^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(zz^*) - zz^*\big\rangle\big| \\ &\leq \sum_{a,b\in\Sigma} p(a)p(b)\big|\big\langle x_a x_b^*, \Phi(x_a x_b^*) - x_a x_b^*\big\rangle\big| \leq \varepsilon. \end{aligned} \tag{3.335}$$

Using the expression of the fidelity function when one of its arguments has rank equal to one, as given by Proposition 3.13, it follows that

$$\mathrm{F}\big(\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(zz^*), zz^*\big) \geq \sqrt{1-\varepsilon}. \tag{3.336}$$

Therefore, by one of the Fuchs–van de Graaf inequalities (Theorem 3.36), it follows that

$$\begin{aligned} &\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(zz^*) - zz^* \big\|_1 \\ &\quad \leq 2\sqrt{1 - \mathrm{F}\big(\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(zz^*), zz^*\big)^2} \leq 2\sqrt{\varepsilon}. \end{aligned} \tag{3.337}$$

Because $\Phi - \mathbb{1}_{\mathrm{L}(\mathcal{X})}$ is a Hermiticity preserving map, the theorem follows by Theorem 3.53. □

3.3.4 Properties of the completely bounded trace norm

This section discusses additional facts concerning the completely bounded trace norm. A few alternative characterizations of the completely bounded trace norm are presented, along with a theorem concerning the completely bounded trace norm of maps having bounded Choi rank.

The maximum output fidelity between completely positive maps

It is possible to characterize the completely bounded trace norm of a map in terms of the *maximum output fidelity* between two completely positive maps derived from the given map. The maximum output fidelity is defined as follows.

Definition 3.59. Let $\Psi_0, \Psi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Y})$ be completely positive maps, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. The *maximum output fidelity* between Ψ_0 and Ψ_1 is defined as

$$\mathrm{F}_{\max}(\Psi_0, \Psi_1) = \max_{\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})} \mathrm{F}\big(\Psi_0(\rho_0), \Psi_1(\rho_1)\big). \tag{3.338}$$

For any choice of vectors of the form $u, v \in \mathcal{X} \otimes \mathcal{Y}$, for $\mathcal{X}$ and $\mathcal{Y}$ being arbitrary complex Euclidean spaces, Corollary 3.24 states that

$$\big\| \mathrm{Tr}_{\mathcal{Y}}\big(vu^*\big) \big\|_1 = \mathrm{F}\big(\mathrm{Tr}_{\mathcal{X}}\big(uu^*\big), \mathrm{Tr}_{\mathcal{X}}\big(vv^*\big)\big). \tag{3.339}$$

It is an extension of this fact that provides the link between the maximum output fidelity and the completely bounded trace norm. In considering this extension, it is convenient to isolate the fact represented by the lemma that follows.

Lemma 3.60. *Let $A_0, A_1 \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be operators, for $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces, and let $\Psi_0, \Psi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Z})$ and $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be the maps defined as*

$$\begin{aligned} \Psi_0(X) &= \mathrm{Tr}_{\mathcal{Y}}\big(A_0 X A_0^*\big), \\ \Psi_1(X) &= \mathrm{Tr}_{\mathcal{Y}}\big(A_1 X A_1^*\big), \end{aligned} \tag{3.340}$$

and

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(A_0 X A_1^*\big), \tag{3.341}$$

for every operator $X \in \mathrm{L}(\mathcal{X})$. Also let $\mathcal{W}$ be a complex Euclidean space and let $u_0, u_1 \in \mathcal{X} \otimes \mathcal{W}$ be vectors. It holds that

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big(u_0 u_1^*\big) \big\|_1 = \mathrm{F}\big(\Psi_0\big(\mathrm{Tr}_{\mathcal{W}}\big(u_0 u_0^*\big)\big), \Psi_1\big(\mathrm{Tr}_{\mathcal{W}}\big(u_1 u_1^*\big)\big)\big). \tag{3.342}$$

Proof. Let $W \in \mathrm{U}(\mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{W}, \mathcal{Z} \otimes \mathcal{Y} \otimes \mathcal{W})$ be the operator defined by the equation

$$W(y \otimes z \otimes w) = z \otimes y \otimes w, \tag{3.343}$$

holding for all $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, and $w \in \mathcal{W}$. In other words, W represents a reordering of tensor factors, from $\mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{W}$ to $\mathcal{Z} \otimes \mathcal{Y} \otimes \mathcal{W}$. It is evident that one has

$$\begin{aligned} \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big(u_0 u_1^*\big) &= \mathrm{Tr}_{\mathcal{Z}}\Big(\big(A_0 \otimes \mathbb{1}_{\mathcal{W}}\big) u_0 u_1^* \big(A_1^* \otimes \mathbb{1}_{\mathcal{W}}\big)\Big) \\ &= \mathrm{Tr}_{\mathcal{Z}}\Big(W\big(A_0 \otimes \mathbb{1}_{\mathcal{W}}\big) u_0 u_1^* \big(A_1^* \otimes \mathbb{1}_{\mathcal{W}}\big) W^*\Big). \end{aligned} \tag{3.344}$$

Applying Corollary 3.24, one has

$$\begin{aligned} &\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big(u_0 u_1^*\big) \big\|_1 \\ &\quad = \mathrm{F}\Big(\mathrm{Tr}_{\mathcal{Y}\otimes\mathcal{W}}\big(W\big(A_0 \otimes \mathbb{1}_{\mathcal{W}}\big) u_0 u_0^* \big(A_0^* \otimes \mathbb{1}_{\mathcal{W}}\big) W^*\big), \\ &\qquad\qquad \mathrm{Tr}_{\mathcal{Y}\otimes\mathcal{W}}\big(W\big(A_1 \otimes \mathbb{1}_{\mathcal{W}}\big) u_1 u_1^* \big(A_1^* \otimes \mathbb{1}_{\mathcal{W}}\big) W^*\big)\Big) \\ &\quad = \mathrm{F}\Big(\mathrm{Tr}_{\mathcal{Y}}\big(A_0 \,\mathrm{Tr}_{\mathcal{W}}\big(u_0 u_0^*\big) A_0^*\big), \mathrm{Tr}_{\mathcal{Y}}\big(A_1 \,\mathrm{Tr}_{\mathcal{W}}\big(u_1 u_1^*\big) A_1^*\big)\Big) \\ &\quad = \mathrm{F}\big(\Psi_0\big(\mathrm{Tr}_{\mathcal{W}}\big(u_0 u_0^*\big)\big), \Psi_1\big(\mathrm{Tr}_{\mathcal{W}}\big(u_1 u_1^*\big)\big)\big), \end{aligned} \tag{3.345}$$

as required. □

Theorem 3.61. *Let $A_0, A_1 \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be operators, for $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces, and let $\Psi_0, \Psi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Z})$ and $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be the maps defined as*

$$\begin{aligned} \Psi_0(X) &= \mathrm{Tr}_{\mathcal{Y}}\big(A_0 X A_0^*\big), \\ \Psi_1(X) &= \mathrm{Tr}_{\mathcal{Y}}\big(A_1 X A_1^*\big), \end{aligned} \tag{3.346}$$

and

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(A_0 X A_1^*\big), \tag{3.347}$$

for every operator $X \in \mathrm{L}(\mathcal{X})$. It holds that

$$\big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 = \mathrm{F}_{\max}(\Psi_0, \Psi_1). \tag{3.348}$$

Proof. Let $\mathcal{W}$ be a complex Euclidean space with $\dim(\mathcal{W}) = \dim(\mathcal{X})$. By Proposition 3.47 and Lemma 3.60, one has

$$\begin{aligned} \big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 &= \max_{u_0, u_1 \in \mathcal{S}(\mathcal{X} \otimes \mathcal{W})} \big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big(u_0 u_1^*\big) \big\|_1 \\ &= \max_{u_0, u_1 \in \mathcal{S}(\mathcal{X} \otimes \mathcal{W})} \mathrm{F}\big(\Psi_0\big(\mathrm{Tr}_{\mathcal{W}}\big(u_0 u_0^*\big)\big), \Psi_1\big(\mathrm{Tr}_{\mathcal{W}}\big(u_1 u_1^*\big)\big)\big) \\ &= \max_{\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})} \mathrm{F}\big(\Psi_0(\rho_0), \Psi_1(\rho_1)\big) \\ &= \mathrm{F}_{\max}(\Psi_0, \Psi_1), \end{aligned} \tag{3.349}$$

as required. □

Remark 3.62. The proof of Theorem 3.61 establishes a connection between those choices of density operators $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ achieving the maximal value in the expression

$$\mathrm{F}_{\max}(\Psi_0, \Psi_1) = \max_{\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})} \mathrm{F}\big(\Psi_0(\rho_0), \Psi_1(\rho_1)\big) \tag{3.350}$$

and the choices of vectors $u_0, u_1 \in \mathcal{S}(\mathcal{X} \otimes \mathcal{W})$ achieving the maximal value in the expression

$$\big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 = \max_{u_0, u_1 \in \mathcal{S}(\mathcal{X} \otimes \mathcal{W})} \big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big(u_0 u_1^*\big) \big\|_1. \tag{3.351}$$

Specifically, for any choice of unit vectors $u_0, u_1 \in \mathcal{S}(\mathcal{X} \otimes \mathcal{W})$, one may take

$$\rho_0 = \mathrm{Tr}_{\mathcal{W}}\big(u_0 u_0^*\big) \quad \text{and} \quad \rho_1 = \mathrm{Tr}_{\mathcal{W}}\big(u_1 u_1^*\big), \tag{3.352}$$

and conversely, for any choice of density operators $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$, one may take $u_0, u_1 \in \mathcal{S}(\mathcal{X} \otimes \mathcal{W})$ to be arbitrary purifications of ρ_0, ρ_1, respectively, with equal values being obtained in the above expressions in both cases.

By combining Theorem 3.61 with the multiplicativity of the completely bounded trace norm with respect to tensor products (Theorem 3.51), one finds that the maximum output fidelity is also multiplicative with respect to tensor products.

Corollary 3.63. *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0, \Psi_0 \in \mathrm{CP}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1, \Psi_1 \in \mathrm{CP}(\mathcal{X}_1, \mathcal{Y}_1)$ be completely positive maps. It holds that*

$$\mathrm{F}_{\max}(\Phi_0 \otimes \Phi_1, \Psi_0 \otimes \Psi_1) = \mathrm{F}_{\max}(\Phi_0, \Psi_0)\, \mathrm{F}_{\max}(\Phi_1, \Psi_1). \tag{3.353}$$

This corollary states a fact that is simple but not necessarily obvious: the maximum output fidelity between two completely positive product maps is achieved for product state inputs. It may be contrasted with some other quantities of interest (such as the minimum output entropy of a quantum channel, to be discussed in Chapter 7) that fail to respect tensor products in this way.

A semidefinite program for maximum output fidelity

It is natural to ask if the value $|||\Phi|||_1$ of the completely bounded trace norm of a given map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ can be efficiently calculated. While there is no closed-form expression that is known to represent this value, it is equal to the optimal value of a semidefinite program that has a simple description in terms of the mapping Φ. In particular, when Theorem 3.61 is combined with the semidefinite program for the fidelity function discussed in Section 3.1.2, a semidefinite program for the completely bounded trace norm is obtained. This allows for an efficient calculation of the value $|||\Phi|||_1$ using a computer, as well as an efficient method of verification through the use of semidefinite programming duality.

In greater detail, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and assume that a Stinespring representation of Φ is known:

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\bigl(A_0 X A_1^*\bigr) \tag{3.354}$$

for all operators $X \in \mathrm{L}(\mathcal{X})$, for $\mathcal{Z}$ being a complex Euclidean space and $A_0, A_1 \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ being operators. Define completely positive maps $\Psi_0, \Psi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Z})$ as

$$\Psi_0(X) = \mathrm{Tr}_{\mathcal{Y}}\bigl(A_0 X A_0^*\bigr) \quad \text{and} \quad \Psi_1(X) = \mathrm{Tr}_{\mathcal{Y}}\bigl(A_1 X A_1^*\bigr) \tag{3.355}$$

for all $X \in \mathrm{L}(\mathcal{X})$, and consider the semidefinite program whose primal problem is as follows:

$$\underline{\text{Primal problem}}$$

$$\begin{aligned} \text{maximize:} \quad & \frac{1}{2}\operatorname{Tr}(Y) + \frac{1}{2}\operatorname{Tr}(Y^*) \\ \text{subject to:} \quad & \begin{pmatrix} \Psi_0(\rho_0) & Y \\ Y^* & \Psi_1(\rho_1) \end{pmatrix} \geq 0 \\ & \rho_0, \rho_1 \in \mathrm{D}(\mathcal{X}),\ Y \in \mathrm{L}(\mathcal{Z}). \end{aligned}$$

Such a semidefinite program may be expressed with greater formality, with respect to the definition of semidefinite programs presented in Section 1.2.2, in the following way. First, one defines a Hermiticity-preserving map

$$\Xi : \mathrm{L}(\mathcal{X} \oplus \mathcal{X} \oplus \mathcal{Z} \oplus \mathcal{Z}) \to \mathrm{L}(\mathbb{C} \oplus \mathbb{C} \oplus \mathcal{Z} \oplus \mathcal{Z}) \tag{3.356}$$

as

$$\begin{aligned} &\Xi \begin{pmatrix} X_0 & \cdot & \cdot & \cdot \\ \cdot & X_1 & \cdot & \cdot \\ \cdot & \cdot & Z_0 & \cdot \\ \cdot & \cdot & \cdot & Z_1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} \operatorname{Tr}(X_0) & 0 & 0 & 0 \\ 0 & \operatorname{Tr}(X_1) & 0 & 0 \\ 0 & 0 & Z_0 - \Psi_0(X_0) & 0 \\ 0 & 0 & 0 & Z_1 - \Psi_1(X_1) \end{pmatrix} \end{aligned} \tag{3.357}$$

for all $X_0, X_1 \in \mathrm{L}(\mathcal{X})$ and $Z_0, Z_1 \in \mathrm{L}(\mathcal{Z})$, and where the dots represent operators on appropriately chosen spaces upon which Ξ does not depend. Then, one may define Hermitian operators $A \in \operatorname{Herm}(\mathcal{X} \oplus \mathcal{X} \oplus \mathcal{Z} \oplus \mathcal{Z})$ and $B \in \operatorname{Herm}(\mathbb{C} \oplus \mathbb{C} \oplus \mathcal{Z} \oplus \mathcal{Z})$ as

$$A = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbb{1} \\ 0 & 0 & \mathbb{1} & 0 \end{pmatrix} \quad \text{and} \quad B = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.358}$$

It is evident that the primal problem specified above is equivalent to the maximization of the quantity $\langle A, X \rangle$ over all choices of

$$X = \begin{pmatrix} X_0 & \cdot & \cdot & \cdot \\ \cdot & X_1 & \cdot & \cdot \\ \cdot & \cdot & Z_0 & Y \\ \cdot & \cdot & Y^* & Z_1 \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X} \oplus \mathcal{Z} \oplus \mathcal{Z}) \tag{3.359}$$

obeying the constraint $\Xi(X) = B$.

The adjoint mapping to Ξ is given by

$$\begin{aligned} &\Xi^* \begin{pmatrix} \lambda_0 & \cdot & \cdot & \cdot \\ \cdot & \lambda_1 & \cdot & \cdot \\ \cdot & \cdot & Z_0 & \cdot \\ \cdot & \cdot & \cdot & Z_1 \end{pmatrix} \\ &\qquad = \frac{1}{2} \begin{pmatrix} \lambda_0 \mathbb{1}_{\mathcal{X}} - \Psi_0^*(Z_0) & 0 & 0 & 0 \\ 0 & \lambda_1 \mathbb{1}_{\mathcal{X}} - \Psi_1^*(Z_1) & 0 & 0 \\ 0 & 0 & Z_0 & 0 \\ 0 & 0 & 0 & Z_1 \end{pmatrix}, \end{aligned} \tag{3.360}$$

so the dual problem corresponding to the semidefinite program (Ξ, A, B) is to minimize the quantity $(\lambda_0 + \lambda_1)/2$ subject to the conditions

$$\lambda_0 \mathbb{1}_{\mathcal{X}} \geq \Psi_0^*(Z_0) \quad \text{and} \quad \lambda_1 \mathbb{1}_{\mathcal{X}} \geq \Psi_1^*(Z_1), \tag{3.361}$$

for $Z_0, Z_1 \in \mathrm{Herm}(\mathcal{Z})$ being Hermitian operators satisfying

$$\begin{pmatrix} Z_0 & 0 \\ 0 & Z_1 \end{pmatrix} \geq \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}. \tag{3.362}$$

Observing that Z_0 and Z_1 must be positive definite in order for (3.362) to be satisfied, along with the fact that Ψ_0^* and Ψ_1^* are positive, one obtains the following statement of the dual problem:

$$\underline{\text{Dual problem}}$$

$$\begin{aligned} \text{minimize:} \quad & \frac{1}{2} \left\| \Psi_0^*(Z_0) \right\| + \frac{1}{2} \left\| \Psi_1^*(Z_1) \right\| \\ \text{subject to:} \quad & \begin{pmatrix} Z_0 & -\mathbb{1}_{\mathcal{Z}} \\ -\mathbb{1}_{\mathcal{Z}} & Z_1 \end{pmatrix} \geq 0 \\ & Z_0, Z_1 \in \mathrm{Pd}(\mathcal{Z}). \end{aligned}$$

To prove that strong duality holds for this semidefinite program, one may observe that the primal problem is feasible and the dual problem is strictly feasible. With respect to the formal specification of the semidefinite program just described, one has that the operator

$$\begin{pmatrix} \rho_0 & 0 & 0 & 0 \\ 0 & \rho_1 & 0 & 0 \\ 0 & 0 & \Psi_0(\rho_0) & 0 \\ 0 & 0 & 0 & \Psi_1(\rho_1) \end{pmatrix} \tag{3.363}$$

is primal feasible, for an arbitrary choice of density operators $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$. One also has that the dual problem is strictly feasible: the operator

$$\begin{pmatrix} \lambda_0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \mathbb{1}_{\mathcal{Z}} & 0 \\ 0 & 0 & 0 & \mathbb{1}_{\mathcal{Z}} \end{pmatrix} \tag{3.364}$$

is strictly dual feasible provided that $\lambda_0 > \|\Psi_0^*(\mathbb{1}_{\mathcal{Z}})\|$ and $\lambda_1 > \|\Psi_1^*(\mathbb{1}_{\mathcal{Z}})\|$. It follows by Slater's theorem (Theorem 1.13) that the primal and dual optimal values are equal, and moreover the primal optimal value is achieved for some choice of a primal feasible operator.

The fact that the optimal value of the semidefinite program is in agreement with the completely bounded norm $|||\Phi|||_1$ follows from Theorem 3.61 together with Theorem 3.17.

It may be noted that the dual problem stated above may be further simplified as follows:

Dual problem (simplified)

$$\begin{aligned} \text{minimize:} \quad & \frac{1}{2}\|\Psi_0^*(Z)\| + \frac{1}{2}\|\Psi_1^*(Z^{-1})\| \\ \text{subject to:} \quad & Z \in \mathrm{Pd}(\mathcal{Y}). \end{aligned}$$

To verify that this problem has the same optimal value as the dual problem stated above, one may first observe that

$$\begin{pmatrix} Z_0 & -\mathbb{1} \\ -\mathbb{1} & Z_1 \end{pmatrix} \tag{3.365}$$

is positive semidefinite if and only if Z_0 and Z_1 are both positive definite and satisfy $Z_1 \geq Z_0^{-1}$. For any such choice of Z_0 and Z_1, the inequality

$$\big\| \Psi_1^*(Z_1) \big\| \geq \big\| \Psi_1^*(Z_0^{-1}) \big\| \tag{3.366}$$

holds by the positivity of Ψ_1^*, implying that there is no loss of generality in restricting one's attention to operators $Z_0 = Z$ and $Z_1 = Z^{-1}$ for $Z \in \mathrm{Pd}(\mathcal{Z})$.

The observation that the simplified dual problem above has optimal value $\|\!|\Phi|\!\|_1$ may be stated in the form of a theorem as follows.

Theorem 3.64. *Let $A_0, A_1 \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be operators, for $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces, and let $\Psi_0, \Psi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Z})$ and $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be the maps defined as*

$$\begin{aligned} \Psi_0(X) &= \mathrm{Tr}_{\mathcal{Y}}\big(A_0 X A_0^*\big), \\ \Psi_1(X) &= \mathrm{Tr}_{\mathcal{Y}}\big(A_1 X A_1^*\big), \end{aligned} \tag{3.367}$$

and

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(A_0 X A_1^*\big), \tag{3.368}$$

for every operator $X \in \mathrm{L}(\mathcal{X})$. It holds that

$$\|\!|\Phi|\!\|_1 = \inf_{Z \in \mathrm{Pd}(\mathcal{Z})} \left(\frac{1}{2} \big\| \Psi_0^*(Z) \big\| + \frac{1}{2} \big\| \Psi_1^*(Z^{-1}) \big\| \right). \tag{3.369}$$

Spectral norm characterization of the completely bounded trace norm

Consider a map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$. One has, by Theorem 2.22, that a given complex Euclidean space $\mathcal{Z}$ admits a Stinespring representation

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(A_0 X A_1^*\big) \tag{3.370}$$

of Φ, for some choice of operators $A_0, A_1 \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$, if and only if the dimension of $\mathcal{Z}$ is at least as large as the Choi rank of Φ. An equivalent condition to (3.370) holding for all operators $X \in \mathrm{L}(\mathcal{X})$ is that

$$J(\Phi) = \mathrm{Tr}_{\mathcal{Z}}\big(\mathrm{vec}(A_0)\,\mathrm{vec}(A_1)^*\big). \tag{3.371}$$

As the next theorem states, the completely bounded trace norm of Φ is equal to the infimum value of the product $\|A_0\|\,\|A_1\|$, ranging over all such choices of A_0 and A_1.

Theorem 3.65 (Smith). *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map, let $\mathcal{Z}$ be a complex Euclidean space satisfying $\dim(\mathcal{Z}) \geq \operatorname{rank}(J(\Phi))$, and let*

$$\begin{aligned}\mathcal{K}_\Phi = \big\{(A_0, A_1) \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z}) \times \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z}) \,:\, \\ J(\Phi) = \operatorname{Tr}_{\mathcal{Z}}\big(\operatorname{vec}(A_0)\operatorname{vec}(A_1)^*\big)\big\}.\end{aligned} \tag{3.372}$$

It holds that

$$\big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 = \inf_{(A_0,A_1)\in\mathcal{K}_\Phi} \|A_0\|\,\|A_1\|. \tag{3.373}$$

Proof. There exists a pair of unit vectors $u, v \in \mathcal{X} \otimes \mathcal{X}$ such that, for any pair of operators $(A_0, A_1) \in \mathcal{K}_\Phi$, one has

$$\big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 = \big\|\operatorname{Tr}_{\mathcal{Z}}\big((A_0 \otimes \mathbb{1}_{\mathcal{X}})uv^*(A_1 \otimes \mathbb{1}_{\mathcal{X}})^*\big)\big\|_1. \tag{3.374}$$

By the monotonicity of the trace norm under partial tracing (1.177) and the multiplicativity of the spectral norm with respect to tensor products, it follows that

$$\begin{aligned}\big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 &\leq \big\|(A_0 \otimes \mathbb{1}_{\mathcal{X}})uv^*(A_1 \otimes \mathbb{1}_{\mathcal{X}})^*\big\|_1 \\ &= \big\|(A_0 \otimes \mathbb{1}_{\mathcal{X}})u\big\|\,\big\|(A_1 \otimes \mathbb{1}_{\mathcal{X}})v\big\| \\ &\leq \|A_0 \otimes \mathbb{1}_{\mathcal{X}}\|\,\|A_1 \otimes \mathbb{1}_{\mathcal{X}}\| \\ &= \|A_0\|\,\|A_1\|.\end{aligned} \tag{3.375}$$

As this inequality holds for every pair $(A_0, A_1) \in \mathcal{K}_\Phi$, it follows that

$$\big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 \leq \inf_{(A_0,A_1)\in\mathcal{K}_\Phi} \|A_0\|\,\|A_1\|. \tag{3.376}$$

It remains to prove the reverse inequality. To this end, fix any pair of operators $(B_0, B_1) \in \mathcal{K}_\Phi$, and define $\Psi_0, \Psi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Z})$ as

$$\Psi_0(X) = \operatorname{Tr}_{\mathcal{Y}}\big(B_0 X B_0^*\big) \quad \text{and} \quad \Psi_1(X) = \operatorname{Tr}_{\mathcal{Y}}\big(B_1 X B_1^*\big) \tag{3.377}$$

for all $X \in \mathrm{L}(\mathcal{X})$, so that

$$\Psi_0^*(Z) = B_0^*(\mathbb{1}_{\mathcal{Y}} \otimes Z)B_0 \quad \text{and} \quad \Psi_1^*(Z) = B_1^*(\mathbb{1}_{\mathcal{Y}} \otimes Z)B_1 \tag{3.378}$$

for every $Z \in \mathrm{L}(\mathcal{Z})$. By Theorem 3.64, the expression (3.369) holds. For any choice of a positive real number $\varepsilon > 0$, there must therefore exist a positive definite operator $Z \in \mathrm{Pd}(\mathcal{Z})$ so that

$$\frac{1}{2}\big\|\Psi_0^*(Z)\big\| + \frac{1}{2}\big\|\Psi_1^*(Z^{-1})\big\| < \big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 + \varepsilon. \tag{3.379}$$

By the arithmetic-geometric mean inequality, it follows that

$$\sqrt{\|\Psi_0^*(Z)\|}\sqrt{\|\Psi_1^*(Z^{-1})\|} < \left|\!\left|\!\left|\Phi\right|\!\right|\!\right|_1 + \varepsilon. \tag{3.380}$$

Setting

$$A_0 = \left(\mathbb{1}_{\mathcal{Y}} \otimes Z^{\frac{1}{2}}\right) B_0 \quad \text{and} \quad A_1 = \left(\mathbb{1}_{\mathcal{Y}} \otimes Z^{-\frac{1}{2}}\right) B_1, \tag{3.381}$$

one has that $(A_0, A_1) \in \mathcal{K}_\Phi$ by the cyclic property of the trace. Moreover, it holds that

$$\begin{aligned} \|A_0\|\|A_1\| &= \sqrt{\|A_0^* A_0\|}\sqrt{\|A_1^* A_1\|} \\ &= \sqrt{\|\Psi_0^*(Z)\|}\sqrt{\|\Psi_1^*(Z^{-1})\|} < \left|\!\left|\!\left|\Phi\right|\!\right|\!\right|_1 + \varepsilon. \end{aligned} \tag{3.382}$$

As it has been established that, for any choice of $\varepsilon > 0$, there exists a pair of operators $(A_0, A_1) \in \mathcal{K}_\Phi$ satisfying the inequality (3.382), it follows that

$$\inf_{(A_0, A_1) \in \mathcal{K}_\Phi} \|A_0\|\|A_1\| \leq \left|\!\left|\!\left|\Phi\right|\!\right|\!\right|_1, \tag{3.383}$$

which completes the proof. □

The completely bounded trace norm of maps with bounded Choi rank

For a given map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ and a complex Euclidean space $\mathcal{Z}$, it holds (by Theorem 3.49) that

$$\left\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\right\|_1 \leq \left|\!\left|\!\left|\Phi\right|\!\right|\!\right|_1, \tag{3.384}$$

with equality under the condition that $\dim(\mathcal{Z}) \geq \dim(\mathcal{X})$. If it is the case that $\dim(\mathcal{Z}) < \dim(\mathcal{X})$, then equality may fail to hold. For instance, the transpose map $\mathrm{T}(X) = X^{\mathsf{T}}$ on an arbitrary complex Euclidean space $\mathcal{X}$ is such that

$$\left\|\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\right\|_1 = \min\bigl\{\dim(\mathcal{X}), \dim(\mathcal{Z})\bigr\} \tag{3.385}$$

for every complex Euclidean space $\mathcal{Z}$.

It is the case, however, that equality holds in (3.384) under a different and generally incomparable assumption, which is that the dimension of $\mathcal{Z}$ is at least as large as the Choi rank of Φ, as the following theorem states.

Theorem 3.66 (Timoney). *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map, and assume $\dim(\mathcal{Z}) \geq \operatorname{rank}(J(\Phi))$. It holds that*

$$\left|\!\left|\!\left|\Phi\right|\!\right|\!\right|_1 = \left\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\right\|_1. \tag{3.386}$$

The proof of Theorem 3.66 to be presented below makes use of the following lemma.

Lemma 3.67. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a positive map, and let $P \in \operatorname{Pos}(\mathcal{Y})$ be a nonzero positive semidefinite operator satisfying $P = \Phi(\rho)$ for some choice of a density operator $\rho \in \mathrm{D}(\mathcal{X})$. There exists a density operator $\sigma \in \mathrm{D}(\mathcal{X})$ with $\operatorname{rank}(\sigma) \leq \operatorname{rank}(P)$ that satisfies $P = \Phi(\sigma)$.*

Proof. Define a set

$$\mathcal{C} = \big\{\xi \in \mathrm{D}(\mathcal{X}) \,:\, \Phi(\xi) = P\big\}. \tag{3.387}$$

The set $\mathcal{C}$ is nonempty by the assumptions of the lemma, and it is evidently both compact and convex. There must therefore exist an extreme point of $\mathcal{C}$. Let σ be such an extreme point and let $r = \operatorname{rank}(\sigma)$. It will be proved that $r \leq \operatorname{rank}(P)$, which suffices to prove the lemma.

Let $n = \dim(\mathcal{X})$ and $m = \operatorname{rank}(P)$, and let $\Pi = \Pi_{\operatorname{im}(P)}$. Define a linear mapping $\Psi : \operatorname{Herm}(\mathcal{X}) \to \operatorname{Herm}(\mathcal{Y} \oplus \mathbb{C})$ as

$$\Psi(H) = \begin{pmatrix} \Pi\Phi(H)\Pi & 0 \\ 0 & \langle \mathbb{1}_{\mathcal{Y}} - \Pi, \Phi(H)\rangle \end{pmatrix} \tag{3.388}$$

for all $H \in \operatorname{Herm}(\mathcal{X})$. The image of Ψ has dimension at most $m^2 + 1$, and therefore the kernel of Ψ is a subspace of $\operatorname{Herm}(\mathcal{X})$ having dimension at least $n^2 - m^2 - 1$. Also define a subspace $\mathcal{W} \subseteq \operatorname{Herm}(\mathcal{X})$ as

$$\mathcal{W} = \big\{H \in \operatorname{Herm}(\mathcal{X}) \,:\, \operatorname{im}(H) \subseteq \operatorname{im}(\sigma) \text{ and } \operatorname{Tr}(H) = 0\big\}. \tag{3.389}$$

The dimension of $\mathcal{W}$ is equal to $r^2 - 1$.

Now consider any operator $H \in \ker(\Psi) \cap \mathcal{W}$. As $\operatorname{im}(H) \subseteq \operatorname{im}(\sigma)$ and σ is positive semidefinite, there must exist a positive real number $\varepsilon > 0$ for which $\sigma + \varepsilon H$ and $\sigma - \varepsilon H$ are both positive semidefinite. As H is traceless, it follow that $\sigma + \varepsilon H$ and $\sigma - \varepsilon H$ are density operators. By the assumption that $H \in \ker(\Psi)$, one has $\langle \mathbb{1}_{\mathcal{Y}} - \Pi, \Phi(H)\rangle = 0$, and therefore

$$\langle \mathbb{1}_{\mathcal{Y}} - \Pi, \Phi(\sigma + \varepsilon H)\rangle = \langle \mathbb{1}_{\mathcal{Y}} - \Pi, P + \varepsilon\Phi(H)\rangle = 0. \tag{3.390}$$

By the positivity of Φ, it follows that

$$\Phi(\sigma + \varepsilon H) = \Pi\Phi(\sigma + \varepsilon H)\Pi = P + \varepsilon\,\Pi\Phi(H)\Pi = P. \tag{3.391}$$

By similar reasoning, $\Phi(\sigma - \varepsilon H) = P$. It has therefore been proved that $\sigma + \varepsilon H$ and $\sigma - \varepsilon H$ are both elements of $\mathcal{C}$; but given that σ was chosen to be an extreme point of $\mathcal{C}$ and

$$\frac{1}{2}\big(\sigma + \varepsilon H\big) + \frac{1}{2}\big(\sigma - \varepsilon H\big) = \sigma, \tag{3.392}$$

it follows that $H = 0$. Consequently, the subspace $\ker(\Psi) \cap \mathcal{W}$ must have dimension 0.

Finally, given that $\mathrm{Herm}(\mathcal{X})$ has dimension n^2, $\ker(\Psi) \subseteq \mathrm{Herm}(\mathcal{X})$ has dimension at least $n^2 - m^2 - 1$, $\mathcal{W} \subseteq \mathrm{Herm}(\mathcal{X})$ has dimension $r^2 - 1$, and $\ker(\Psi) \cap \mathcal{W}$ has dimension 0, it follows that

$$\big(n^2 - m^2 - 1\big) + \big(r^2 - 1\big) \leq n^2, \tag{3.393}$$

and therefore

$$r^2 \leq m^2 + 2. \tag{3.394}$$

As r and m are positive integers, it follows that $r \leq m$, which completes the proof. □

Proof of Theorem 3.66. One may choose operators $A_0, A_1 \in \mathrm{L}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ such that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(A_0 X A_1^*\big) \tag{3.395}$$

for all $X \in \mathrm{L}(\mathcal{X})$, by Theorem 2.22. By Theorem 3.61, it follows that

$$\big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1 = \mathrm{F}_{\max}(\Psi_0, \Psi_1) \tag{3.396}$$

for $\Psi_0, \Psi_1 \in \mathrm{CP}(\mathcal{X}, \mathcal{Z})$ being the completely positive maps defined by

$$\Psi_0(X) = \mathrm{Tr}_{\mathcal{Y}}\big(A_0 X A_0^*\big) \quad \text{and} \quad \Psi_1(X) = \mathrm{Tr}_{\mathcal{Y}}\big(A_1 X A_1^*\big) \tag{3.397}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be density operators that satisfy

$$\mathrm{F}\big(\Psi_0(\rho_0), \Psi_1(\rho_1)\big) = \mathrm{F}_{\max}(\Psi_0, \Psi_1) = \big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1. \tag{3.398}$$

The operators $P_0 = \Psi_0(\rho_0)$ and $P_1 = \Psi_1(\rho_1)$ are elements of $\mathrm{Pos}(\mathcal{Z})$, and therefore their rank cannot exceed the dimension of $\mathcal{Z}$. It follows from

Lemma 3.67 that there exist density operators $\sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X})$, whose rank also does not exceed the dimension of $\mathcal{Z}$, for which it holds that

$$\Psi_0(\sigma_0) = P_0 \quad \text{and} \quad \Psi_1(\sigma_1) = P_1. \tag{3.399}$$

Therefore, one has that

$$\mathrm{F}\big(\Psi_0(\sigma_0), \Psi_1(\sigma_1)\big) = \big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1. \tag{3.400}$$

Because σ_0 and σ_1 have rank at most the dimension of $\mathcal{Z}$, there must exist unit vectors $u_0, u_1 \in \mathcal{X} \otimes \mathcal{Z}$ satisfying

$$\sigma_0 = \mathrm{Tr}_{\mathcal{Z}}\big(u_0 u_0^*\big) \quad \text{and} \quad \sigma_1 = \mathrm{Tr}_{\mathcal{Z}}\big(u_1 u_1^*\big). \tag{3.401}$$

By Lemma 3.60, one has that

$$\big\| \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(u_0 u_1^*) \big\|_1 = \mathrm{F}\big(\Psi_0(\sigma_0), \Psi_1(\sigma_1)\big) = \big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1, \tag{3.402}$$

which establishes that

$$\big\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big\|_1 \geq \big|\!\big|\!\big|\Phi\big|\!\big|\!\big|_1. \tag{3.403}$$

As the reverse inequality holds by Theorem 3.49, the proof is complete. □

Corollary 3.68. *Let $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be channels, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and let $\mathcal{Z}$ be any complex Euclidean space with*

$$\dim(\mathcal{Z}) \geq 2\,\mathrm{rank}(J(\Phi_0 - \Phi_1)), \tag{3.404}$$

There exists a unit vector $u \in \mathcal{X} \otimes \mathcal{Z}$ such that

$$\big\| \big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(uu^*) - \big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(uu^*) \big\|_1 = \big|\!\big|\!\big|\Phi_0 - \Phi_1\big|\!\big|\!\big|_1. \tag{3.405}$$

Proof. The theorem is vacuous when $\Phi_0 = \Phi_1$, so it will be assumed that this is not the case. Let $\mathcal{W}$ be a complex Euclidean space having dimension equal to $\mathrm{rank}(J(\Phi_0 - \Phi_1))$. By Theorem 3.66, it holds that

$$\big|\!\big|\!\big|\Phi_0 - \Phi_1\big|\!\big|\!\big|_1 = \big\|\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})} - \Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big\|_1 \tag{3.406}$$

By Lemma 3.52, it follows that there exists a unit vector $v \in \mathcal{X} \otimes \mathcal{W} \otimes \mathcal{V}$, for $\mathcal{V}$ being any complex Euclidean space with dimension equal to 2, such that

$$\big\| \big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W}\otimes\mathcal{V})}\big)(vv^*) - \big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W}\otimes\mathcal{V})}\big)(vv^*) \big\|_1 \geq \big|\!\big|\!\big|\Phi_0 - \Phi_1\big|\!\big|\!\big|_1. \tag{3.407}$$

Now, under the assumption that $\dim(\mathcal{Z}) \geq 2\,\mathrm{rank}(J(\Phi_0 - \Phi_1))$, there must exist a linear isometry of the form $V \in \mathrm{U}(\mathcal{W} \otimes \mathcal{V}, \mathcal{Z})$. One may set

$$u = (\mathbb{1}_{\mathcal{X}} \otimes V)v \tag{3.408}$$

to obtain

$$\begin{aligned}
&\big\| \big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(uu^*) - \big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(uu^*) \big\|_1 \\
&\qquad = \big\| (\mathbb{1}_{\mathcal{Y}} \otimes V)\big(\big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W}\otimes\mathcal{V})}\big)(vv^*) \\
&\qquad\qquad - \big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W}\otimes\mathcal{V})}\big)(vv^*)\big)(\mathbb{1}_{\mathcal{Y}} \otimes V^*)\big\|_1 \\
&\qquad = \big\| \big(\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W}\otimes\mathcal{V})}\big)(vv^*) - \big(\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W}\otimes\mathcal{V})}\big)(vv^*) \big\|_1 \\
&\qquad \geq \big\| \Phi_0 - \Phi_1 \big\|_1
\end{aligned} \tag{3.409}$$

by the isometric invariance of the trace norm together with (3.407). As the reverse inequality holds for all unit vectors $u \in \mathcal{X} \otimes \mathcal{Z}$ by Theorem 3.49, the proof is complete. □

3.4 Exercises

3.1. Let $\mathcal{X}$ be a complex Euclidean space, let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be states, let $\varepsilon > 0$ be a positive real number, and assume that there exists a measurement $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X})$ for which it holds that

$$\frac{1}{2}\langle \mu(0), \rho_0 \rangle + \frac{1}{2}\langle \mu(1), \rho_1 \rangle \geq \varepsilon. \tag{3.410}$$

Prove that, for every positive integer n, there exists a measurement

$$\nu : \{0,1\} \to \mathrm{Pos}(\mathcal{X}^{\otimes n}) \tag{3.411}$$

such that

$$\frac{1}{2}\langle \nu(0), \rho_0^{\otimes n} \rangle + \frac{1}{2}\langle \nu(1), \rho_1^{\otimes n} \rangle \geq 1 - \frac{1}{2}\exp\left(-\frac{n\varepsilon^2}{8}\right). \tag{3.412}$$

3.2. Suppose $\mathcal{X}$ and $\mathcal{Y}$ are complex Euclidean spaces, $P, Q \in \mathrm{Pos}(\mathcal{X})$ are positive semidefinite operators, and $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ is a trace-preserving and positive (but not necessarily completely positive) map. Prove that

$$\mathrm{F}(P, Q) \leq \mathrm{F}(\Phi(P), \Phi(Q)). \tag{3.413}$$

3.3. Find an example of two channels $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, for some choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, such that

$$\big\| \Phi_0(\rho) - \Phi_1(\rho) \big\|_1 < \big\| \Phi_0 - \Phi_1 \big\|_1 \tag{3.414}$$

for every density operator $\rho \in \mathrm{D}(\mathcal{X})$.

3.4. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a map. Prove that

$$\big|\!\big|\!\big| \Phi \big|\!\big|\!\big|_1 = \max_{\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})} \big\| \big(\mathbb{1}_{\mathcal{Y}} \otimes \sqrt{\rho_0}\big) J(\Phi) \big(\mathbb{1}_{\mathcal{Y}} \otimes \sqrt{\rho_1}\big) \big\|_1. \tag{3.415}$$

3.5. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces with $\dim(\mathcal{X}) = n$ and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$. Prove that

$$|\!|\!| \Phi |\!|\!|_1 \leq \| J(\Phi) \|_1 \leq n |\!|\!| \Phi |\!|\!|_1. \tag{3.416}$$

3.6. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $H \in \mathrm{Herm}(\mathcal{Y} \otimes \mathcal{X})$ be a Hermitian operator. Consider the problem of maximizing the value

$$\langle H, J(\Psi) \rangle \tag{3.417}$$

over all choices of a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$. Prove that

$$\langle H, J(\Phi) \rangle = \sup\{ \langle H, J(\Psi) \rangle \,:\, \Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y}) \} \tag{3.418}$$

if and only if the operator $\mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi))$ is Hermitian and satisfies

$$\mathbb{1}_{\mathcal{Y}} \otimes \mathrm{Tr}_{\mathcal{Y}}(HJ(\Phi)) \geq H. \tag{3.419}$$

3.5 Bibliographic remarks

The task of quantum state discrimination was evidently first formulated (in abstract terms) by Helstrom [100], although instances of the task with respect to specific quantum physical systems had certainly been considered earlier. Theorem 3.4 was proved by Helstrom [100] for the restricted case of projective measurements, and by Holevo [104] for general measurements.

Theorem 3.7 was proved by Gutoski and Watrous [84], and a slightly weaker form of the theorem (for finite sets of states) was proved by Jain [124] around the same time, using Sion's minmax theorem. Jain's proof

extends easily to the more general case, and this is the proof that has been presented in this chapter.

Theorem 3.9 is attributed to Holevo [104, 105] and Yuen, Kennedy, and Lax [234, 235]. The semidefinite programming formulation that has been used in the proof of this theorem is due to Yuen, Kennedy, and Lax, although it was not recognized as a semidefinite program in their work (as their work predates much of the development of semidefinite programming). Eldar, Megretski, and Verghese [70] recognized this optimization problem as a semidefinite program. The pretty good measurement was so-named and popularized by Hausladen and Wootters [93]—it is one among a family of measurements introduced earler by Belavkin [26] and considered in other works (such as Eldar and Forney [69]). Theorem 3.10 is due to Barnum and Knill [20].

The fidelity function was introduced by Uhlmann [209], who referred to it as the *transition probability*. (Uhlmann defined the transition probability as the square of the fidelity function, as it has been defined in this book. Many authors follow the convention of referring to the square of the fidelity function as the fidelity function.) Uhlmann also proved Theorem 3.23 and observed several elementary properties of the fidelity function in the same paper. Corollary 3.21 is due to Alberti [5], and Theorem 3.30 is due to Alberti and Uhlmann [7]. The term *fidelity* was first introduced by Jozsa [127], who presented a simplified proof of Uhlmann's theorem.

A variant of Corollary 3.15 was proved by Winter in [230], stated in terms of the trace distance rather than the fidelity. Theorem 3.27 is due to Fuchs and Caves [74], Theorem 3.32 is due to Spekkens and Rudolph [196], and Theorem 3.36 is due to Fuchs and van de Graaf [75]. Theorem 3.17 and the semidefinite program associated with that theorem was independently found by Killoran [128] and Watrous [223].

The channel fidelity was introduced by Schumacher [186], who named it the entanglement fidelity and established some basic results about it, including a derivation of its expression as represented by Proposition 3.34 and the fact that it is invariant under the choice of the purification used to define it. A proof of Proposition 3.35 appears in Nielsen [166].

The relevance of the completely bounded trace norm to the theory of quantum information and computation appears to have first been realized by Kitaev [129], who took the spectral norm characterization (Theorem 3.65) as the definition and proved its equivalence to Definition 3.46. Several

basic properties of the completely bounded trace norm, including those summarized in Proposition 3.47, appear in the same paper, as well as in the work of Aharonov, Kitaev, and Nisan [3]. Kitaev used the notation $\|\cdot\|_\diamond$ rather than $|||\cdot|||_1$ when referring to the completely bounded trace norm, which led to its being referred to as the "diamond norm." This norm's close relationship to the completely bounded norm used in the study of operator algebras later came to be realized; in finite dimensions, the norm known as the completely bounded norm is essentially equivalent to the completely bounded trace norm, with the definition being based on the spectral norm rather than the trace norm. The book of Paulsen [168] provides an overview of the properties, uses, and history of this norm in the subject of operator algebras.

Example 3.39 appears to be somewhat of a folklore result. A variation of this example appears in Kretschmann, Schlingemann, and Werner [136], and had been recognized by others (including this author) a couple of years earlier. A different example having a similar character, but not giving as sharp a separation, appears in Kitaev, Shen, and Vyalyi [130]. (See Example 11.1 and the text immediately following in that book.) Underlying all of these examples is the observation that the transpose mapping provides a separation between the induced trace norm and the completely bounded trace norm; an equivalent example goes back (at least) to Arveson [14].

Theorem 3.42 is equivalent to a theorem of Russo and Dye [179]. Results similar to Lemma 3.48 appear in Gilchrist, Langford, and Nielsen [78] and Watrous [219], and a fact equivalent to Lemma 3.52 appears in Rosgen and Watrous [177]. Theorem 3.57 is stated in Aharonov, Kitaev, and Nisan [3] for the case of unitary channels, and an equivalent statement was proved by Childs, Preskill, and Renes [46]. The extension of this statement to isometric channels through the use of the Toeplitz–Hausdorff theorem can reasonably be described as being routine. A similar bound to the one in Theorem 3.58 appears in Kretschmann and Werner [137].

Theorem 3.61 appears (as an exercise, together with a solution) in Kitaev, Shen, and Vyalyi [130]. Theorem 3.65 is due to Smith [195]. Theorem 3.66 is due to Timoney [201], and the proof of this theorem given in this chapter is from Watrous [220].

The completely bounded trace norm can be expressed as a semidefinite program in a few different ways, as was proved by Watrous [222, 223]. Ben-Aroya and Ta-Shma [28] independently proved that the completely

bounded trace norm can be efficiently computed through the use of convex programming techniques, and a similar statement is made for the somewhat simpler task of computing the completely bounded trace norm of the difference between two channels by Gilchrist, Langford, and Nielsen [78]. Other computational methods for evaluating the completely bounded trace norm, but not accompanied by proofs of their computationally efficiency, were devised by Zarikian [236] and Johnston, Kribs, and Paulsen [126].

Chapter 4

Unital channels and majorization

This chapter studies the class of *unital channels*, together with the related notion of *majorization* for Hermitian operators. The following definition of unital channels will be used throughout the chapter.

Definition 4.1. Let $\mathcal{X}$ be a complex Euclidean space. A channel $\Phi \in \mathrm{C}(\mathcal{X})$ is a *unital channel* if and only if $\Phi(\mathbb{1}_{\mathcal{X}}) = \mathbb{1}_{\mathcal{X}}$.[1]

The first section of the chapter introduces various subclasses of unital channels, and the second section investigates properties of unital channels in general. The third section discusses majorization for Hermitian operators, together with an analogous notion for real vectors.

4.1 Subclasses of unital channels

Three subclasses of unital channels are introduced in the subsections that follow: *mixed-unitary channels*, *Weyl-covariant channels*, and *Schur channels*. Various properties of these classes, as well as relationships among them, and to general unital channels, are discussed.

[1] The requirement that unital channels take the form $\Phi \in \mathrm{C}(\mathcal{X})$, for some choice of a complex Euclidean space $\mathcal{X}$, is both natural and convenient with respect to the specific topics to be discussed in this chapter. One could, more generally, consider any channel of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, for some choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, to be a unital channel if the condition $\Phi(\mathbb{1}_{\mathcal{X}}) = \mathbb{1}_{\mathcal{Y}}$ is met. In order for this requirement to be met by a channel, it must hold that $\dim(\mathcal{Y}) = \dim(\mathcal{X})$; and for this reason there is little generality lost in restricting the definition of unital channels to those of the form $\Phi \in \mathrm{C}(\mathcal{X})$.

4.1.1 Mixed-unitary channels

Every unitary channel is evidently unital, as is any convex combination of unitary channels. Channels of the later sort will be referred to as *mixed-unitary* channels, as the following definition makes precise.

Definition 4.2. Let $\mathcal{X}$ be a complex Euclidean space and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a channel. It is said that Φ is a *mixed-unitary channel* if and only if there exists an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and a collection of unitary operators $\{U_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{X})$ such that

$$\Phi(X) = \sum_{a \in \Sigma} p(a) U_a X U_a^* \tag{4.1}$$

for every $X \in \mathrm{L}(\mathcal{X})$. Equivalently, a mapping $\Phi \in \mathrm{C}(\mathcal{X})$ is a mixed-unitary channel if and only if it is a convex combination of unitary channels.

An example of a unital channel that is not mixed-unitary

While every mixed-unitary channel is necessarily unital, the converse of this statement does not hold, as the following example illustrates.

Example 4.3. Let $\mathcal{X} = \mathbb{C}^3$ and define $\Phi \in \mathrm{C}(\mathcal{X})$ as

$$\Phi(X) = \frac{1}{2}\operatorname{Tr}(X)\mathbb{1} - \frac{1}{2}X^{\mathsf{T}} \tag{4.2}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Example 3.39 has established that Φ is a channel, and it is evident that Φ is unital. The channel Φ is, however, not a mixed-unitary channel. To see this, observe first that

$$\Phi(X) = A_1 X A_1^* + A_2 X A_2^* + A_3 X A_3^* \tag{4.3}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for

$$A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{-1}{\sqrt{2}} & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 0 \\ \frac{-1}{\sqrt{2}} & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.4}$$

The fact that the expression (4.3) does indeed hold for all $X \in \mathrm{L}(\mathcal{X})$ follows from the observation that the Choi representation of the map defined by the

right-hand side of that equation is in agreement with $J(\Phi)$, as calculated in Example 3.39:

$$\frac{1}{2}\mathbb{1}\otimes\mathbb{1} - \frac{1}{2}W = \sum_{k=1}^{3} \operatorname{vec}(A_k)\operatorname{vec}(A_k)^*. \tag{4.5}$$

The collection $\{A_j^* A_k : 1 \leq j,k \leq 3\}$ includes the following operators:

$$A_1^*A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}, \quad A_1^*A_2 = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_1^*A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{-1}{2} & 0 & 0 \end{pmatrix},$$

$$A_2^*A_1 = \begin{pmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_2^*A_2 = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}, \quad A_2^*A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}, \tag{4.6}$$

$$A_3^*A_1 = \begin{pmatrix} 0 & 0 & \frac{-1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad A_3^*A_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}, \quad A_3^*A_3 = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This is a linearly independent collection, as an inspection reveals. It follows from Theorem 2.31 that Φ is an extreme point of the set of channels $\mathrm{C}(\mathcal{X})$. As Φ is not itself a unitary channel, it follows that it cannot be expressed as a convex combination of unitary channels.

Pinching channels

Many interesting examples of mixed-unitary channels are known. One type of channel, called a *pinching channel*, provides a collection of examples.

Definition 4.4. Let $\mathcal{X}$ be a complex Euclidean space. A channel $\Phi \in \mathrm{C}(\mathcal{X})$ is said to be a *pinching channel*, or simply a *pinching*, if there exists a collection $\{\Pi_a : a \in \Sigma\}$ of projection operators satisfying

$$\sum_{a\in\Sigma} \Pi_a = \mathbb{1}_{\mathcal{X}} \tag{4.7}$$

(i.e., such that the set $\{\Pi_a : a \in \Sigma\}$ represents a projective measurement) for which

$$\Phi(X) = \sum_{a\in\Sigma} \Pi_a X \Pi_a \tag{4.8}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

The action of the channel defined by (4.8) on a register X is equivalent to X being measured with respect to a nondestructive measurement defined by $\{\Pi_a : a \in \Sigma\}$, followed by the measurement outcome being discarded.

Example 4.5. The channel $\Phi \in \mathrm{C}(\mathbb{C}^5)$ defined as

$$\Phi(X) = \Pi_0 X \Pi_0 + \Pi_1 X \Pi_1 \tag{4.9}$$

for

$$\Pi_0 = E_{1,1} + E_{2,2} \qquad \text{and} \qquad \Pi_1 = E_{3,3} + E_{4,4} + E_{5,5} \tag{4.10}$$

is an example of a pinching channel. This channel has the following action on a general operator in $\mathrm{L}(\mathcal{X})$, expressed in matrix form:

$$\Phi \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} & \alpha_{1,4} & \alpha_{1,5} \\ \alpha_{2,1} & \alpha_{2,2} & \alpha_{2,3} & \alpha_{2,4} & \alpha_{2,5} \\ \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} & \alpha_{3,4} & \alpha_{3,5} \\ \alpha_{4,1} & \alpha_{4,2} & \alpha_{4,3} & \alpha_{4,4} & \alpha_{4,5} \\ \alpha_{5,1} & \alpha_{5,2} & \alpha_{5,3} & \alpha_{5,4} & \alpha_{5,5} \end{pmatrix} = \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & 0 & 0 & 0 \\ \alpha_{2,1} & \alpha_{2,2} & 0 & 0 & 0 \\ 0 & 0 & \alpha_{3,3} & \alpha_{3,4} & \alpha_{3,5} \\ 0 & 0 & \alpha_{4,3} & \alpha_{4,4} & \alpha_{4,5} \\ 0 & 0 & \alpha_{5,3} & \alpha_{5,4} & \alpha_{5,5} \end{pmatrix}. \tag{4.11}$$

The action of this channel is suggestive of the matrix representing the input operator being "pinched," causing a certain pattern of off-diagonal entries to become 0, which explains the terminology used to describe such maps. When a pinching channel is defined by a collection of projection operators that are not diagonal in the standard basis, the term is not descriptive in this way, but it is used nevertheless.

While it is not immediate from the definition that every pinching channel is a mixed-unitary channel, it is fairly straightforward to establish that this is so, as the proof of the following proposition reveals.

Proposition 4.6. *Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, and let $\{\Pi_a : a \in \Sigma\}$ be a collection of projection operators on $\mathcal{X}$ satisfying*

$$\sum_{a \in \Sigma} \Pi_a = \mathbb{1}_{\mathcal{X}}. \tag{4.12}$$

The channel $\Phi \in \mathrm{C}(\mathcal{X})$ defined by

$$\Phi(X) = \sum_{a \in \Sigma} \Pi_a X \Pi_a \tag{4.13}$$

for all $X \in \mathrm{L}(\mathcal{X})$ is a mixed-unitary channel.

Proof. Consider the collection $\{-1,1\}^{\Sigma}$ containing those vectors in $\mathcal{X}$ having entries drawn from the set $\{-1,1\}$, and define a unitary operator

$$U_w = \sum_{a\in\Sigma} w(a)\Pi_a \tag{4.14}$$

for every such vector $w \in \{-1,1\}^{\Sigma}$. It holds that

$$\frac{1}{2^{|\Sigma|}} \sum_{w\in\{-1,1\}^{\Sigma}} U_w X U_w^* = \frac{1}{2^{|\Sigma|}} \sum_{a,b\in\Sigma} \sum_{w\in\{-1,1\}^{\Sigma}} w(a)w(b)\Pi_a X \Pi_b \tag{4.15}$$

for every $X \in \mathrm{L}(\mathcal{X})$. To simplify this expression, one may observe that

$$\frac{1}{2^{|\Sigma|}} \sum_{w\in\{-1,1\}^{\Sigma}} w(a)w(b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases} \tag{4.16}$$

for every choice of $a, b \in \Sigma$, and therefore

$$\frac{1}{2^{|\Sigma|}} \sum_{w\in\{-1,1\}^{\Sigma}} U_w X U_w^* = \sum_{a\in\Sigma} \Pi_a X \Pi_a = \Phi(X) \tag{4.17}$$

for every $X \in \mathrm{L}(\mathcal{X})$. This demonstrates that Φ is a mixed-unitary channel, as required. □

Example 4.7. The completely dephasing channel $\Delta \in \mathrm{C}(\mathcal{X})$ defined on any complex Euclidean space $\mathcal{X} = \mathbb{C}^{\Sigma}$ is an example of a pinching channel, as it is defined according to Definition 4.4 by the collection of projection operators $\{E_{a,a} : a \in \Sigma\}$. By Proposition 4.6, it follows that Δ is a mixed-unitary channel.

Environment-assisted channel correction

Mixed-unitary channels have an alternative characterization based on the notion of *environment-assisted channel correction*. Let $\Phi \in \mathrm{C}(\mathcal{X})$ be a channel, represented in Stinespring form as

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \tag{4.18}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for some choice of a complex Euclidean space $\mathcal{Z}$ and an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{X}\otimes\mathcal{Z})$. Environment-assisted channel correction refers to the existence of an alphabet Σ, a collection of channels

$$\{\Psi_a : a \in \Sigma\} \subset \mathrm{C}(\mathcal{X}), \tag{4.19}$$

and a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{Z})$, for which the equation

$$X = \sum_{a \in \Sigma} \Psi_a\bigl(\mathrm{Tr}_{\mathcal{Z}}\bigl((\mathbb{1}_{\mathcal{X}} \otimes \mu(a))AXA^*\bigr)\bigr) \tag{4.20}$$

holds for all $X \in \mathrm{L}(\mathcal{X})$.

An interpretation of the equation (4.20) is as follows. One imagines that a register X contains a quantum state $\rho \in \mathrm{D}(\mathcal{X})$. The action of the mapping $X \mapsto AXA^*$ has the effect of encoding this state into the state of the pair (X, Z), for Z being a second register. By discarding the register Z, the register X is left in the state $\Phi(\rho)$, which may potentially be quite different from ρ. In essence, the register Z represents an "environment," to which some part of the encoding of ρ may have escaped or leaked. The measurement μ on Z, followed by the application of Ψ_a to X (for whichever outcome $a \in \Sigma$ resulted from the measurement), is viewed as an attempt to *correct* X, so that it is transformed back into ρ. The equation (4.20) represents the situation in which a perfect correction of this sort is accomplished.

The following theorem implies that a perfect correction of the sort just described is possible if and only if Φ is a mixed-unitary channel.

Theorem 4.8. *Let $\mathcal{X}$ and $\mathcal{Z}$ be complex Euclidean spaces, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{X} \otimes \mathcal{Z})$ be an isometry, and let $\Phi \in \mathrm{C}(\mathcal{X})$ be the channel defined by*

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\bigl(AXA^*\bigr) \tag{4.21}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The following two are statements equivalent:

1. *Φ is a mixed-unitary channel.*
2. *There exists an alphabet Σ, a collection of channels $\{\Psi_a : a \in \Sigma\} \subset \mathrm{C}(\mathcal{X})$, and a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{Z})$ for which*

$$X = \sum_{a \in \Sigma} \Psi_a\bigl(\mathrm{Tr}_{\mathcal{Z}}\bigl((\mathbb{1}_{\mathcal{X}} \otimes \mu(a))AXA^*\bigr)\bigr) \tag{4.22}$$

   for all $X \in \mathrm{L}(\mathcal{X})$.

Proof. Assume first that statement 1 holds, so that

$$\Phi(X) = \sum_{a \in \Sigma} p(a) U_a X U_a^* \tag{4.23}$$

for every $X \in \mathrm{L}(\mathcal{X})$, for some choice of an alphabet Σ, a collection of unitary operators $\{U_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{X})$, and a probability vector $p \in \mathcal{P}(\Sigma)$. There

is no loss of generality in assuming $|\Sigma| \geq \dim(\mathcal{Z})$; one may add any finite number of elements to Σ, take $p(a) = 0$, and choose $U_a \in \mathrm{U}(\mathcal{X})$ arbitrarily for the added elements, maintaining the validity of the expression (4.23). By this assumption, there must exist a collection $\{v_a : a \in \Sigma\} \subset \mathcal{Z}$ of vectors for which

$$\sum_{a \in \Sigma} v_a v_a^* = \mathbb{1}_{\mathcal{Z}}. \tag{4.24}$$

Fix such collection, and define operators $\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$ as

$$A_a = (\mathbb{1}_{\mathcal{X}} \otimes v_a^*) A \tag{4.25}$$

for each $a \in \Sigma$. It holds that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(AXA^*\big) = \sum_{a \in \Sigma} A_a X A_a^* \tag{4.26}$$

for every $X \in \mathrm{L}(\mathcal{X})$. Therefore, by Corollary 2.23, there must exist a unitary operator $W \in \mathrm{U}(\mathbb{C}^{\Sigma})$ such that

$$\sqrt{p(a)}\, U_a = \sum_{b \in \Sigma} W(a,b) A_b \tag{4.27}$$

for every $a \in \Sigma$.

For each symbol $a \in \Sigma$, define a vector $u_a \in \mathcal{Z}$ as

$$u_a = \sum_{b \in \Sigma} \overline{W(a,b)} v_b, \tag{4.28}$$

and define $\mu : \Sigma \to \mathrm{Pos}(\mathcal{Z})$ as $\mu(a) = u_a u_a^*$ for each $a \in \Sigma$. Because W is a unitary operator, it holds that

$$\sum_{a \in \Sigma} \mu(a) = \sum_{a,b,c \in \Sigma} \overline{W(a,b)} W(a,c) v_b v_c^* = \sum_{b \in \Sigma} v_b v_b^* = \mathbb{1}_{\mathcal{Z}}, \tag{4.29}$$

and therefore μ is a measurement. Also define a collection $\{\Psi_a : a \in \Sigma\}$ of channels as

$$\Psi_a(X) = U_a^* X U_a \tag{4.30}$$

for every $X \in \mathrm{L}(\mathcal{X})$ and $a \in \Sigma$.

Now, it holds that

$$(\mathbb{1}_{\mathcal{X}} \otimes u_a^*) A = \sum_{b \in \Sigma} W(a,b) A_b = \sqrt{p(a)}\, U_a, \tag{4.31}$$

and therefore

$$\mathrm{Tr}_{\mathcal{Z}}\big((\mathbb{1}_{\mathcal{X}} \otimes \mu(a))AXA^*\big) = p(a)U_a X U_a^*, \tag{4.32}$$

for each $a \in \Sigma$. It follows that

$$\sum_{a\in\Sigma} \Psi_a\big(\mathrm{Tr}_{\mathcal{Z}}\big((\mathbb{1}_{\mathcal{X}} \otimes \mu(a))AXA^*\big)\big) = \sum_{a\in\Sigma} p(a)U_a^* U_a X U_a^* U_a = X \tag{4.33}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Statement 1 therefore implies statement 2.

Next, assume statement 2 holds. For each $a \in \Sigma$, define $\Phi_a \in \mathrm{CP}(\mathcal{X})$ as

$$\Phi_a(X) = \mathrm{Tr}_{\mathcal{Z}}\big((\mathbb{1}_{\mathcal{X}} \otimes \mu(a))AXA^*\big) \tag{4.34}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Also let

$$\{A_{a,b} : a \in \Sigma,\, b \in \Gamma\} \quad \text{and} \quad \{B_{a,b} : a \in \Sigma,\, b \in \Gamma\} \tag{4.35}$$

be collections of operators in $\mathrm{L}(\mathcal{X})$, for a suitable choice of an alphabet Γ, yielding Kraus representations

$$\Psi_a(X) = \sum_{b\in\Gamma} A_{a,b} X A_{a,b}^* \quad \text{and} \quad \Phi_a(X) = \sum_{c\in\Gamma} B_{a,c} X B_{a,c}^* \tag{4.36}$$

for all $a \in \Sigma$ and $X \in \mathrm{L}(\mathcal{X})$. (Taking a common alphabet Γ as an index set for these representations is only done to simplify notation and causes no loss of generality; one is free to include the zero operator among the Kraus operators of either map any number of times.) By the assumption that statement 2 holds, one has

$$\sum_{a\in\Sigma} \Psi_a \Phi_a = \mathbb{1}_{\mathrm{L}(\mathcal{X})}, \tag{4.37}$$

and therefore the Choi representations of the two sides equation (4.37) must agree:

$$\sum_{a\in\Sigma} \sum_{b,c\in\Gamma} \mathrm{vec}(A_{a,b}B_{a,c})\,\mathrm{vec}(A_{a,b}B_{a,c})^* = \mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*. \tag{4.38}$$

There must therefore exist a collection $\{\alpha_{a,b,c} : a \in \Sigma,\, b, c \in \Gamma\}$ of complex numbers for which the equation

$$A_{a,b}B_{a,c} = \alpha_{a,b,c}\mathbb{1}_{\mathcal{X}} \tag{4.39}$$

holds for all $a \in \Sigma$ and $b, c \in \Gamma$. This collection must also evidently satisfy the constraint

$$\sum_{a \in \Sigma} \sum_{b,c \in \Gamma} |\alpha_{a,b,c}|^2 = 1. \tag{4.40}$$

Consequently, one has

$$\sum_{b \in \Gamma} |\alpha_{a,b,c}|^2 \mathbb{1}_{\mathcal{X}} = \sum_{b \in \Gamma} B_{a,c}^* A_{a,b}^* A_{a,b} B_{a,c} = B_{a,c}^* B_{a,c} \tag{4.41}$$

for every $a \in \Sigma$ and $c \in \Gamma$, owing to the fact that each mapping Ψ_a is a channel. For every $a \in \Sigma$ and $c \in \Gamma$ it must therefore hold that

$$B_{a,c} = \beta_{a,c} U_{a,c} \tag{4.42}$$

for some choice of a unitary operator $U_{a,c} \in \mathrm{U}(\mathcal{X})$ and a complex number $\beta_{a,c} \in \mathbb{C}$ satisfying

$$|\beta_{a,c}|^2 = \sum_{b \in \Gamma} |\alpha_{a,b,c}|^2. \tag{4.43}$$

It follows that

$$\Phi(X) = \sum_{a \in \Sigma} \Phi_a(X) = \sum_{a \in \Sigma} \sum_{c \in \Gamma} p(a,c) U_{a,c} X U_{a,c}^*, \tag{4.44}$$

for $p \in \mathcal{P}(\Sigma \times \Gamma)$ being the probability vector defined as

$$p(a,c) = |\beta_{a,c}|^2 \tag{4.45}$$

for each $a \in \Sigma$ and $c \in \Gamma$. The channel Φ is therefore mixed-unitary, so it has been proved that statement 2 implies statement 1. □

Mixed-unitary channels and Carathéodory's theorem

Every mixed-unitary channel $\Phi \in \mathrm{C}(\mathcal{X})$ is, by definition, an element of the convex hull of the set of unitary channels. Using Carathéodory's theorem (Theorem 1.9), one may obtain upper-bounds on the number of unitary channels that must be averaged to obtain any mixed-unitary channel. The following proposition proves one bound along these lines.

Proposition 4.9. *Let $\mathcal{X}$ be a complex Euclidean space, let $n = \dim(\mathcal{X})$, and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a mixed-unitary channel. There exists a positive integer m satisfying*

$$m \leq n^4 - 2n^2 + 2, \tag{4.46}$$

a collection of unitary operators $\{U_1, \ldots, U_m\} \subset \mathrm{U}(\mathcal{X})$*, and a probability vector* $(p_1, \ldots, p_m)$ *such that*

$$\Phi(X) = \sum_{k=1}^{m} p_k U_k X U_k^* \tag{4.47}$$

for all $X \in \mathrm{L}(\mathcal{X})$*.*

Proof. Consider the linear map $\Xi : \mathrm{Herm}(\mathcal{X} \otimes \mathcal{X}) \to \mathrm{Herm}(\mathcal{X} \oplus \mathcal{X})$ defined by the equation

$$\Xi(X \otimes Y) = \begin{pmatrix} \mathrm{Tr}(X)Y & 0 \\ 0 & \mathrm{Tr}(Y)X \end{pmatrix} \tag{4.48}$$

for all $X, Y \in \mathrm{Herm}(\mathcal{X})$, and fix any orthogonal basis

$$\bigl\{\mathbb{1}, H_1, \ldots, H_{n^2-1}\bigr\} \tag{4.49}$$

of $\mathrm{Herm}(\mathcal{X})$ containing the identity operator. It holds that

$$\Xi(H_j \otimes H_k) = 0 \tag{4.50}$$

for every choice of $j, k \in \{1, \ldots, n^2 - 1\}$, while the operators

$$\Xi(\mathbb{1} \otimes H_k), \quad \Xi(H_k \otimes \mathbb{1}), \quad \text{and} \quad \Xi(\mathbb{1} \otimes \mathbb{1}), \tag{4.51}$$

ranging over all choices of $k \in \{1, \ldots, n^2 - 1\}$, are all nonzero and pairwise orthogonal. The kernel of Ξ is therefore equal to the subspace spanned by the orthogonal collection

$$\bigl\{H_j \otimes H_k \,:\, 1 \leq j, k \leq n^2 - 1\bigr\}. \tag{4.52}$$

In particular, the dimension of the kernel of the mapping Ξ is

$$(n^2 - 1)^2 = n^4 - 2n^2 + 1. \tag{4.53}$$

Next, consider any unitary operator $U \in \mathrm{U}(\mathcal{X})$, and let $\Psi_U \in \mathrm{C}(\mathcal{X})$ be the unitary channel defined as

$$\Psi_U(X) = UXU^* \tag{4.54}$$

for every $X \in \mathrm{L}(\mathcal{X})$. Evaluating the mapping Ξ defined above on the Choi representation of Ψ_U yields

$$\Xi(J(\Psi_U)) = \Xi(\mathrm{vec}(U)\,\mathrm{vec}(U)^*) = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix}. \tag{4.55}$$

The Choi representation of Ψ_U is therefore drawn from an affine subspace of $\mathrm{Herm}(\mathcal{X} \otimes \mathcal{X})$ having dimension $n^4 - 2n^2 + 1$.

Because Φ is a mixed-unitary channel, the Choi representation $J(\Phi)$ of Φ is equal to a convex combination of operators having the form $J(\Psi_U)$, for U ranging over a finite set of unitary operators. It therefore follows from Carathéodory's theorem that

$$J(\Phi) = \sum_{k=1}^{m} p_k J\bigl(\Psi_{U_k}\bigr) \tag{4.56}$$

for some choice of a positive integer

$$m \leq n^4 - 2n^2 + 2, \tag{4.57}$$

unitary operators $U_1, \ldots, U_m \in \mathrm{U}(\mathcal{X})$, and a probability vector $(p_1, \ldots, p_m)$. Equivalently,

$$\Phi(X) = \sum_{k=1}^{m} p_k U_k X U_k^* \tag{4.58}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for the same choice of m, $U_1, \ldots, U_m$, and $(p_1, \ldots, p_m)$, which completes the proof. □

A similar technique to the one used in the proof above may be used to obtain an upper bound on the number of channels, drawn from an arbitrary collection, that must be averaged to obtain a given element in the convex hull of that collection. As a corollary, one obtains a different bound (which is almost always better than the one from the previous proposition) on the number of unitary channels that must be averaged to obtain a given mixed-unitary channel.

Theorem 4.10. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\mathcal{A} \subseteq \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be any nonempty collection of channels, and let $\Phi \in \mathrm{conv}(\mathcal{A})$ be a channel in the convex hull of $\mathcal{A}$. There exists a positive integer*

$$m \leq \mathrm{rank}(J(\Phi))^2, \tag{4.59}$$

a probability vector $(p_1, \ldots, p_m)$, and a selection of channels $\Psi_1, \ldots, \Psi_m \in \mathcal{A}$ such that

$$\Phi = p_1 \Psi_1 + \cdots + p_m \Psi_m. \tag{4.60}$$

Proof. Let $r = \operatorname{rank}(J(\Phi))$ and take Π to be the projection operator onto the image of $J(\Phi)$. Define a linear map

$$\Xi : \operatorname{Herm}(\mathcal{Y} \otimes \mathcal{X}) \to \operatorname{Herm}(\mathbb{C} \oplus (\mathcal{Y} \otimes \mathcal{X}) \oplus (\mathcal{Y} \otimes \mathcal{X})) \tag{4.61}$$

as

$$\Xi(H) = \begin{pmatrix} \operatorname{Tr}(H) & 0 & 0 \\ 0 & (\mathbb{1} - \Pi)H(\mathbb{1} - \Pi) & (\mathbb{1} - \Pi)H\Pi \\ 0 & \Pi H(\mathbb{1} - \Pi) & 0 \end{pmatrix} \tag{4.62}$$

for each $H \in \operatorname{Herm}(\mathcal{Y} \otimes \mathcal{X})$. It holds that $\Xi(H) = 0$ for precisely those Hermitian operators H satisfying

$$H = \Pi H \Pi \quad \text{and} \quad \operatorname{Tr}(H) = 0, \tag{4.63}$$

and therefore the kernel of Ξ has dimension $r^2 - 1$.

Let

$$\mathcal{B} = \{\Psi \in \mathcal{A} : \operatorname{im}(J(\Psi)) \subseteq \operatorname{im}(J(\Phi))\}, \tag{4.64}$$

and observe that $\Phi \in \operatorname{conv}(\mathcal{B})$, by virtue of the fact that $\Phi \in \operatorname{conv}(\mathcal{A})$. For each channel $\Psi \in \mathcal{B}$ it holds that

$$\Xi(J(\Psi)) = \begin{pmatrix} \dim(\mathcal{X}) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.65}$$

There is therefore an affine subspace of $\operatorname{Herm}(\mathcal{Y} \otimes \mathcal{X})$ of dimension $r^2 - 1$ that contains $J(\Psi)$, for every $\Psi \in \mathcal{B}$. As $J(\Phi)$ is a convex combination of operators in this affine subspace, it follows from Carathéodory's theorem that there exists an integer $m \leq (r^2 - 1) + 1 = r^2$, a selection of channels $\Psi_1, \ldots, \Psi_m \in \mathcal{B} \subseteq \mathcal{A}$, and a probability vector $(p_1, \ldots, p_m)$ such that

$$J(\Phi) = p_1 J(\Psi_1) + \cdots + p_m J(\Psi_m). \tag{4.66}$$

The equation (4.66) is equivalent to (4.60), which completes the proof. □

Corollary 4.11. *Let $\mathcal{X}$ be a complex Euclidean space and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a mixed-unitary channel. There exists a positive integer $m \leq \operatorname{rank}(J(\Phi))^2$, a selection of unitary operators $U_1, \ldots, U_m \in \mathrm{U}(\mathcal{X})$, and a probability vector $(p_1, \ldots, p_m)$ such that*

$$\Phi(X) = \sum_{k=1}^{m} p_k U_k X U_k^* \tag{4.67}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

4.1.2 Weyl-covariant channels

This section concerns the *Weyl-covariant channels*, which are a class of unital channels that relate (in multiple ways) to a collection of operators called the *discrete Weyl operators*.

Discrete Weyl operators

For every positive integer n, the set $\mathbb{Z}_n$ is defined as

$$\mathbb{Z}_n = \{0, \ldots, n-1\}. \tag{4.68}$$

This set forms a ring, with respect to addition and multiplication modulo n, and whenever elements of $\mathbb{Z}_n$ appear in arithmetic expressions in this book, the default assumption is that the operations are to be taken modulo n.

Now, assume that a positive integer n has been fixed, and let $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$. The *discrete Weyl operators* are a collection of unitary operators acting on $\mathcal{X}$, defined in the following way.[2] One first defines a scalar value

$$\zeta = \exp\left(\frac{2\pi i}{n}\right), \tag{4.69}$$

along with unitary operators

$$U = \sum_{c \in \mathbb{Z}_n} E_{c+1,c} \quad \text{and} \quad V = \sum_{c \in \mathbb{Z}_n} \zeta^c E_{c,c}. \tag{4.70}$$

For each pair $(a, b) \in \mathbb{Z}_n \times \mathbb{Z}_n$, the discrete Weyl operator $W_{a,b} \in \mathrm{U}(\mathcal{X})$ is then defined as

$$W_{a,b} = U^a V^b, \tag{4.71}$$

or equivalently as

$$W_{a,b} = \sum_{c \in \mathbb{Z}_n} \zeta^{bc} E_{a+c,c}. \tag{4.72}$$

[2] It is sometimes convenient to extend the definition of the discrete Weyl operators from complex Euclidean spaces of the form $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$ to arbitrary complex Euclidean spaces $\mathcal{X} = \mathbb{C}^{\Sigma}$, simply by placing Σ in correspondence with $\mathbb{Z}_n$, for $n = |\Sigma|$, in some fixed but otherwise arbitrary way.

Example 4.12. For $n = 2$, the discrete Weyl operators (in matrix form) are given by

$$\begin{aligned} W_{0,0} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad W_{0,1} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \\ W_{1,0} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad W_{1,1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{4.73}$$

Equivalently,

$$W_{0,0} = \mathbb{1}, \quad W_{0,1} = \sigma_z, \quad W_{1,0} = \sigma_x, \quad W_{1,1} = -i\sigma_y, \tag{4.74}$$

where

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{4.75}$$

are the *Pauli operators*.

It holds that

$$UV = \sum_{c \in \mathbb{Z}_n} \zeta^c E_{c+1,c} \quad \text{and} \quad VU = \sum_{c \in \mathbb{Z}_n} \zeta^{c+1} E_{c+1,c}, \tag{4.76}$$

from which the commutation relation

$$VU = \zeta UV \tag{4.77}$$

follows. Identities that may be derived using this relation, together with straightforward calculations, include

$$\overline{W_{a,b}} = W_{a,-b}, \quad W_{a,b}^{\mathsf{T}} = \zeta^{-ab} W_{-a,b}, \quad \text{and} \quad W_{a,b}^* = \zeta^{ab} W_{-a,-b} \tag{4.78}$$

for all $a, b \in \mathbb{Z}_n$, and

$$W_{a,b} W_{c,d} = \zeta^{bc} W_{a+c,b+d} = \zeta^{bc-ad} W_{c,d} W_{a,b} \tag{4.79}$$

for all $a, b, c, d \in \mathbb{Z}_n$.

From the equation

$$\sum_{c \in \mathbb{Z}_n} \zeta^{ac} = \begin{cases} n & \text{if } a = 0 \\ 0 & \text{if } a \in \{1, \ldots, n-1\} \end{cases} \tag{4.80}$$

it follows that

$$\mathrm{Tr}(W_{a,b}) = \begin{cases} n & \text{if } (a,b) = (0,0) \\ 0 & \text{otherwise.} \end{cases} \tag{4.81}$$

Combining this observation with (4.79) yields

$$\langle W_{a,b}, W_{c,d} \rangle = \begin{cases} n & \text{if } (a,b) = (c,d) \\ 0 & \text{if } (a,b) \neq (c,d) \end{cases} \tag{4.82}$$

for all $a, b, c, d \in \mathbb{Z}_n$. The set

$$\left\{ \frac{1}{\sqrt{n}} W_{a,b} \,:\, (a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n \right\} \tag{4.83}$$

therefore forms an orthonormal set. Because the cardinality of this set is equal to the dimension of $\mathrm{L}(\mathcal{X})$, it therefore forms an orthonormal basis for this space.

The *discrete Fourier transform operator* $F \in \mathrm{U}(\mathcal{X})$, defined as

$$F = \frac{1}{\sqrt{n}} \sum_{a,b \in \mathbb{Z}_n} \zeta^{ab} E_{a,b}\,, \tag{4.84}$$

has a special connection with the discrete Weyl operators. The fact that F is unitary may be verified by a direct calculation:

$$F^* F = \frac{1}{n} \sum_{a,b,c \in \mathbb{Z}_n} \zeta^{a(b-c)} E_{c,b} = \sum_{b \in \mathbb{Z}_n} E_{b,b} = \mathbb{1}. \tag{4.85}$$

It may also be verified that $FU = VF$ and $FV = U^* F$, from which it follows that

$$F W_{a,b} = \zeta^{-ab} W_{-b,a} F \tag{4.86}$$

for all $a, b \in \mathbb{Z}_n$.

Weyl-covariant maps and channels

A map $\Phi \in \mathrm{T}(\mathcal{X})$, for $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$ as above, is a *Weyl-covariant mapping* if and only if it commutes with the action of conjugation by every discrete Weyl operator, as the following definition makes precise.

Definition 4.13. Let $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$ for n being a positive integer. A map $\Phi \in \mathrm{T}(\mathcal{X})$ is a *Weyl-covariant map* if and only if

$$\Phi\bigl(W_{a,b} X W_{a,b}^*\bigr) = W_{a,b} \Phi(X) W_{a,b}^* \tag{4.87}$$

for every $X \in \mathrm{L}(\mathcal{X})$ and $(a, b) \in \mathbb{Z}_n \times \mathbb{Z}_n$. If, in addition to being a Weyl-covariant map, Φ is a channel, then Φ is said to be a *Weyl-covariant channel*.

From this definition it follows that the set of Weyl-covariant maps of the form $\Phi \in \mathrm{T}(\mathcal{X})$ is a linear subspace of $\mathrm{T}(\mathcal{X})$; for any two Weyl-covariant maps $\Phi, \Psi \in \mathrm{T}(\mathcal{X})$ and scalars $\alpha, \beta \in \mathbb{C}$, the map $\alpha\Phi + \beta\Psi$ is also Weyl-covariant. It follows from this observation that the set of Weyl-covariant channels of the form $\Phi \in \mathrm{C}(\mathcal{X})$ is a convex subset of $\mathrm{C}(\mathcal{X})$.

The next theorem provides two alternative characterizations of Weyl-covariant maps. One characterization states that a map is Weyl-covariant if and only if each discrete Weyl operator is an eigenoperator[3] of that map. The other characterization states that a map is Weyl-covariant if and only if it is a linear combination of conjugations by discrete Weyl operators. The two characterizations are related by the discrete Fourier transform operator.

Theorem 4.14. *Let $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$ for n being a positive integer, and let $\Phi \in \mathrm{T}(\mathcal{X})$ be a map. The following statements are equivalent:*

1. *Φ is a Weyl-covariant map.*
2. *There exists an operator $A \in \mathrm{L}(\mathcal{X})$ such that*

$$\Phi(W_{a,b}) = A(a, b) W_{a,b} \tag{4.88}$$

   for all $(a, b) \in \mathbb{Z}_n \times \mathbb{Z}_n$.
3. *There exists an operator $B \in \mathrm{L}(\mathcal{X})$ such that*

$$\Phi(X) = \sum_{a,b \in \mathbb{Z}_n} B(a, b) W_{a,b} X W_{a,b}^* \tag{4.89}$$

   for all $X \in \mathrm{L}(\mathcal{X})$.

Under the assumption that these three statements hold, the operators A and B in statements 2 and 3 are related by the equation

$$A^{\mathsf{T}} = n F^* B F. \tag{4.90}$$

[3] The term *eigenoperator* should be interpreted in the natural way—as an operator analogue of an eigenvector for a linear map defined on a space of operators.

Proof. Assume Φ is a Weyl-covariant map and consider the operator

$$W_{a,b}^* \Phi(W_{a,b}), \tag{4.91}$$

for $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$ chosen arbitrarily. For every choice of $(c,d) \in \mathbb{Z}_n \times \mathbb{Z}_n$, it holds that

$$\begin{aligned} W_{a,b}^* \Phi(W_{a,b}) W_{c,d}^* &= W_{a,b}^* W_{c,d}^* W_{c,d} \Phi(W_{a,b}) W_{c,d}^* \\ &= W_{a,b}^* W_{c,d}^* \Phi(W_{c,d} W_{a,b} W_{c,d}^*) = W_{c,d}^* W_{a,b}^* \Phi(W_{a,b} W_{c,d} W_{c,d}^*) \\ &= W_{c,d}^* W_{a,b}^* \Phi(W_{a,b}), \end{aligned} \tag{4.92}$$

where the second equality has used the Weyl-covariance of Φ and the third equality has used the fact that

$$W_{c,d} W_{a,b} = \alpha \, W_{a,b} W_{c,d} \quad \text{and} \quad W_{a,b}^* W_{c,d}^* = \overline{\alpha} \, W_{c,d}^* W_{a,b}^* \tag{4.93}$$

for $\alpha = \zeta^{ad-bc}$. It follows that

$$[W_{a,b}^* \Phi(W_{a,b}), W_{c,d}^*] = 0 \tag{4.94}$$

for all $(c,d) \in \mathbb{Z}_n \times \mathbb{Z}_n$. As the set of all discrete Weyl operators forms a basis for $\mathrm{L}(\mathcal{X})$, it must therefore hold that $W_{a,b}^* \Phi(W_{a,b})$ commutes with all operators in $\mathrm{L}(\mathcal{X})$, and is therefore equal to a scalar multiple of the identity operator.

As this is true for every choice of $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$, it follows that one may choose an operator $A \in \mathrm{L}(\mathcal{X})$ so that

$$W_{a,b}^* \Phi(W_{a,b}) = A(a,b) \mathbb{1}, \tag{4.95}$$

and therefore

$$\Phi(W_{a,b}) = A(a,b) W_{a,b}, \tag{4.96}$$

for all $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$. Statement 1 therefore implies statement 2.

The reverse implication, that statement 2 implies statement 1, is implied by the commutation relation (4.79). In greater detail, suppose statement 2 holds, and let $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$. For each pair $(c,d) \in \mathbb{Z}_n \times \mathbb{Z}_n$, one has

$$\begin{aligned} \Phi(W_{a,b} W_{c,d} W_{a,b}^*) &= \zeta^{bc-ad} \Phi(W_{c,d}) = A(c,d) \zeta^{bc-ad} W_{c,d} \\ &= A(c,d) W_{a,b} W_{c,d} W_{a,b}^* = W_{a,b} \Phi(W_{c,d}) W_{a,b}^*, \end{aligned} \tag{4.97}$$

and therefore, again using the fact that the discrete Weyl operators form a basis for $\mathrm{L}(\mathcal{X})$, one has

$$\Phi(W_{a,b} X W_{a,b}^*) = W_{a,b} \Phi(X) W_{a,b}^* \tag{4.98}$$

for all $X \in \mathrm{L}(\mathcal{X})$ by linearity.

Now assume statement 3 holds for some choice of $B \in \mathrm{L}(\mathcal{X})$. Using the commutation relation (4.79), it follows that

$$\Phi(W_{c,d}) = \sum_{a,b \in \mathbb{Z}_n} B(a,b) W_{a,b} W_{c,d} W_{a,b}^* = \sum_{a,b \in \mathbb{Z}_n} \zeta^{bc-ad} B(a,b) W_{c,d} \tag{4.99}$$

for every pair $(c,d) \in \mathbb{Z}_n \times \mathbb{Z}_n$. Choosing $A \in \mathrm{L}(\mathcal{X})$ so that

$$A(c,d) = \sum_{a,b \in \mathbb{Z}_n} \zeta^{bc-ad} B(a,b) \tag{4.100}$$

for all $(c,d) \in \mathbb{Z}_n \times \mathbb{Z}_n$, which is equivalent to $A = (nF^* B F)^{\mathsf{T}}$, one has that

$$\Phi(W_{c,d}) = A(c,d) W_{c,d} \tag{4.101}$$

for all $(c,d) \in \mathbb{Z}_n \times \mathbb{Z}_n$. Statement 3 therefore implies statement 2, with the operators A and B being related as claimed.

Finally, assume statement 2 holds for some choice of $A \in \mathrm{L}(\mathcal{X})$, and define $B = \frac{1}{n} F A^{\mathsf{T}} F^*$. By a similar calculation to the one used to establish the previous implication, one has

$$\begin{aligned} \Phi(W_{c,d}) &= A(c,d) W_{c,d} \\ &= \sum_{a,b \in \mathbb{Z}_n} \zeta^{bc-ad} B(a,b) W_{c,d} = \sum_{a,b \in \mathbb{Z}_n} B(a,b) W_{a,b} W_{c,d} W_{a,b}^* \end{aligned} \tag{4.102}$$

for every pair $(c,d) \in \mathbb{Z}_n \times \mathbb{Z}_n$, and therefore

$$\Phi(X) = \sum_{a,b \in \mathbb{Z}_n} B(a,b) W_{a,b} X W_{a,b}^* \tag{4.103}$$

for all $X \in \mathrm{L}(\mathcal{X})$ by linearity. Statement 2 therefore implies statement 3, where again A and B are related as claimed. □

Corollary 4.15. *Let $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$ for n being a positive integer, and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a Weyl-covariant channel. There exists a probability vector $p \in \mathcal{P}(\mathbb{Z}_n \times \mathbb{Z}_n)$ such that*

$$\Phi(X) = \sum_{a,b \in \mathbb{Z}_n} p(a,b) W_{a,b} X W_{a,b}^* \tag{4.104}$$

for all $X \in \mathrm{L}(\mathcal{X})$. In particular, it holds that Φ is a mixed-unitary channel.

Proof. By Theorem 4.14, there exists an operator $B \in \mathrm{L}(\mathcal{X})$ such that

$$\Phi(X) = \sum_{a,b \in \mathbb{Z}_n} B(a,b) W_{a,b} X W_{a,b}^* \tag{4.105}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It follows that

$$J(\Phi) = \sum_{a,b \in \mathbb{Z}_n} B(a,b) \operatorname{vec}(W_{a,b}) \operatorname{vec}(W_{a,b})^*, \tag{4.106}$$

which is a positive semidefinite operator given the assumption that Φ is completely positive. This implies that $B(a,b)$ is nonnegative for every pair $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$, by virtue of the fact that the vectors

$$\{\operatorname{vec}(W_{a,b}) \,:\, a,b \in \mathbb{Z}_n\} \tag{4.107}$$

form an orthogonal set. It holds that

$$\operatorname{Tr}(\Phi(X)) = \sum_{a,b \in \mathbb{Z}_n} B(a,b) \operatorname{Tr}\bigl(W_{a,b} X W_{a,b}^*\bigr) = \sum_{a,b \in \mathbb{Z}_n} B(a,b) \operatorname{Tr}(X) \tag{4.108}$$

for every $X \in \mathrm{L}(\mathcal{X})$, and therefore

$$\sum_{a,b \in \mathbb{Z}_n} B(a,b) = 1 \tag{4.109}$$

by the assumption that Φ preserves trace. Defining $p(a,b) = B(a,b)$ for every pair $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$, one has that p is a probability vector, which completes the proof. □

Completely depolarizing and dephasing channels

The *completely depolarizing channel* $\Omega \in \mathrm{C}(\mathcal{X})$ and the *completely dephasing channel* $\Delta \in \mathrm{C}(\mathcal{X})$ are defined, for any choice of a complex Euclidean space $\mathcal{X} = \mathbb{C}^{\Sigma}$, as follows:

$$\Omega(X) = \frac{\operatorname{Tr}(X)}{\dim(\mathcal{X})} \mathbb{1}_{\mathcal{X}} \quad \text{and} \quad \Delta(X) = \sum_{a \in \Sigma} X(a,a) E_{a,a} \tag{4.110}$$

for all $X \in \mathrm{L}(\mathcal{X})$ (q.v. Section 2.2.3). In the case that the complex Euclidean space $\mathcal{X}$ takes the form $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$ for a positive integer n, these channels are both examples of Weyl-covariant channels.

That the completely depolarizing channel is a Weyl-covariant channel follows from the observation that

$$\Omega(W_{a,b}) = \begin{cases} W_{a,b} & \text{if } (a,b) = (0,0) \\ 0 & \text{if } (a,b) \neq (0,0), \end{cases} \tag{4.111}$$

or equivalently $\Omega(W_{a,b}) = E_{0,0}(a,b)W_{a,b}$, for every $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$. Thus, by Theorem 4.14, together with the observation that

$$\frac{1}{n} F E_{0,0} F^* = \frac{1}{n^2} \sum_{a,b \in \mathbb{Z}_n} E_{a,b}, \tag{4.112}$$

one has that

$$\Omega(X) = \frac{1}{n^2} \sum_{a,b \in \mathbb{Z}_n} W_{a,b} X W_{a,b}^* \tag{4.113}$$

for all $X \in \mathrm{L}(\mathcal{X})$. An alternative way to establish the validity of the equation (4.113) is to observe that the Choi operator of the map defined by the right-hand side of that equation is in agreement with the Choi operator of Ω:

$$\frac{1}{n^2} \sum_{a,b \in \mathbb{Z}_n} \operatorname{vec}(W_{a,b}) \operatorname{vec}(W_{a,b})^* = \frac{1}{n} \mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{X}} = J(\Omega). \tag{4.114}$$

As mentioned in the footnote on page 231, one may translate the notion of a discrete Weyl operator from a space of the form $\mathbb{C}^{\mathbb{Z}_n}$ to an arbitrary complex Euclidean space $\mathbb{C}^{\Sigma}$ through any fixed correspondence between the elements of Σ and $\mathbb{Z}_n$ (assuming $n = |\Sigma|$). It follows that the completely depolarizing channel $\Omega \in \mathrm{C}(\mathcal{X})$ is a mixed-unitary channel for any choice of a complex Euclidean space $\mathcal{X} = \mathbb{C}^{\Sigma}$, as it is equal to the Weyl-covariant channel defined above with respect to any chosen correspondence between Σ and $\mathbb{Z}_n$.

The completely dephasing channel is a Weyl-covariant channel, as is evident from the observation that

$$\Delta(W_{a,b}) = \begin{cases} W_{a,b} & \text{if } a = 0 \\ 0 & \text{if } a \neq 0, \end{cases} \tag{4.115}$$

or equivalently $\Omega(W_{a,b}) = A(a,b)W_{a,b}$ for

$$A = \sum_{c \in \mathbb{Z}_n} E_{0,c}, \tag{4.116}$$

for all $(a, b) \in \mathbb{Z}_n \times \mathbb{Z}_n$. By Theorem 4.14, together with the observation that $FA^{\mathsf{T}}F^* = A$, it follows that

$$\Delta(X) = \frac{1}{n} \sum_{c \in \mathbb{Z}_n} W_{0,c} X W_{0,c}^* \tag{4.117}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

4.1.3 Schur channels

Schur channels, which are defined as follows, represent another interesting sub-class of unital channels.

Definition 4.16. Let $\mathcal{X} = \mathbb{C}^{\Sigma}$ be the complex Euclidean space indexed by a given alphabet Σ. A map $\Phi \in \mathrm{T}(\mathcal{X})$ is said to be a *Schur map* if and only if there exists an operator $A \in \mathrm{L}(\mathcal{X})$ satisfying

$$\Phi(X) = A \odot X, \tag{4.118}$$

where $A \odot X$ denotes the entry-wise product of A and X:

$$(A \odot X)(a, b) = A(a, b)X(a, b) \tag{4.119}$$

for all $a, b \in \Sigma$. If, in addition, the map Φ is a channel, then it is said to be a *Schur channel*.

The following proposition provides a simple condition under which a given Schur map is completely positive (or, equivalently, positive).

Proposition 4.17. *Let Σ be an alphabet, let $\mathcal{X} = \mathbb{C}^{\Sigma}$, let $A \in \mathrm{L}(\mathcal{X})$ be an operator, and let $\Phi \in \mathrm{T}(\mathcal{X})$ be the Schur map defined as*

$$\Phi(X) = A \odot X \tag{4.120}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The following statements are equivalent:

1. *A is positive semidefinite.*
2. *Φ is positive.*
3. *Φ is completely positive.*

Proof. Suppose A is positive semidefinite. It holds that

$$J(\Phi) = \sum_{a,b\in\Sigma} \Phi(E_{a,b}) \otimes E_{a,b} = \sum_{a,b\in\Sigma} A(a,b) E_{a,b} \otimes E_{a,b} = VAV^* \tag{4.121}$$

for $V \in \mathrm{U}(\mathcal{X}, \mathcal{X} \otimes \mathcal{X})$ being the isometry defined as

$$V = \sum_{a\in\Sigma} (e_a \otimes e_a) e_a^*. \tag{4.122}$$

This implies that $J(\Phi)$ is positive semidefinite, so Φ is completely positive by Theorem 2.22. It has been proved that statement 1 implies statement 3.

Statement 3 trivially implies statement 2 as every completely positive map is positive.

Finally, assume that Φ is positive. The operator $X \in \mathrm{L}(\mathcal{X})$ defined as

$$X(a,b) = 1 \tag{4.123}$$

for all $a, b \in \Sigma$ is positive semidefinite. By the positivity of Φ, it therefore holds that $\Phi(X) = A$ is positive semidefinite. Statement 2 therefore implies statement 1, which completes the proof. □

In a similar spirit to the previous proposition, the following proposition provides a simple condition under which a given Schur map preserves trace (or, equivalently, is unital).

Proposition 4.18. *Let Σ be an alphabet, let $\mathcal{X} = \mathbb{C}^\Sigma$, let $A \in \mathrm{L}(\mathcal{X})$ be an operator, and let $\Phi \in \mathrm{T}(\mathcal{X})$ be the Schur mapping defined as*

$$\Phi(X) = A \odot X \tag{4.124}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The following statements are equivalent:

1. *$A(a,a) = 1$ for every $a \in \Sigma$*
2. *Φ is trace-preserving.*
3. *Φ is unital.*

Proof. Suppose $A(a,a) = 1$ for every $a \in \Sigma$. It follows that Φ is unital, as

$$\Phi(\mathbb{1}) = A \odot \mathbb{1} = \sum_{a\in\Sigma} A(a,a) E_{a,a} = \sum_{a\in\Sigma} E_{a,a} = \mathbb{1}. \tag{4.125}$$

It also follows that Φ is trace-preserving, as

$$\begin{aligned} \operatorname{Tr}(\Phi(X)) &= \sum_{a\in\Sigma} (A \odot X)(a,a) \\ &= \sum_{a\in\Sigma} A(a,a)X(a,a) = \sum_{a\in\Sigma} X(a,a) = \operatorname{Tr}(X) \end{aligned} \tag{4.126}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

The assumption that Φ is trace-preserving implies that

$$A(a,a) = \operatorname{Tr}(A(a,a)E_{a,a}) = \operatorname{Tr}(\Phi(E_{a,a})) = \operatorname{Tr}(E_{a,a}) = 1 \tag{4.127}$$

for all $a \in \Sigma$. Statements 1 and 2 are therefore equivalent.

Finally, the assumption that Φ is unital implies

$$\sum_{a\in\Sigma} A(a,a)E_{a,a} = \Phi(\mathbb{1}) = \mathbb{1} = \sum_{a\in\Sigma} E_{a,a}, \tag{4.128}$$

and therefore $A(a,a) = 1$ for every $a \in \Sigma$. Statements 1 and 3 are therefore equivalent. □

Completely positive Schur maps may alternatively be characterized as the class of maps having Kraus representations consisting only of equal pairs of diagonal operators, as the following theorem states.

Theorem 4.19. *Let Σ be an alphabet, let $\mathcal{X} = \mathbb{C}^{\Sigma}$ be the complex Euclidean space indexed by Σ, and let $\Phi \in \mathrm{CP}(\mathcal{X})$ be a completely positive map. The following statements are equivalent:*

1. *Φ is a Schur map.*
2. *There exists a Kraus representation of Φ having the form*

$$\Phi(X) = \sum_{a\in\Gamma} A_a X A_a^*, \tag{4.129}$$

   for some alphabet Γ, such that $A_a \in \mathrm{L}(\mathcal{X})$ is a diagonal operator for each $a \in \Gamma$.
3. *For every Kraus representation of Φ having the form (4.129), A_a is a diagonal operator for each $a \in \Gamma$.*

Proof. Suppose first that Φ is a Schur map, given by

$$\Phi(X) = P \odot X \tag{4.130}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for some operator $P \in \mathrm{L}(\mathcal{X})$. By the assumption that Φ is completely positive, Proposition 4.17 implies that P is positive semidefinite. As was computed in the proof of that proposition, the Choi representation of Φ is given by

$$J(\Phi) = VPV^* \tag{4.131}$$

for

$$V = \sum_{b \in \Sigma} (e_b \otimes e_b) e_b^*. \tag{4.132}$$

Consider an arbitrary Kraus representation of Φ having the form (4.129), for some alphabet Γ and a collection $\{A_a : a \in \Gamma\} \subset \mathrm{L}(\mathcal{X})$ of operators. As the Choi representation of the map defined by the right-hand side of that equation must agree with (4.131), it holds that

$$\sum_{a \in \Gamma} \mathrm{vec}(A_a)\, \mathrm{vec}(A_a)^* = VPV^*, \tag{4.133}$$

and therefore

$$\mathrm{vec}(A_a) \in \mathrm{im}(V) = \mathrm{span}\{e_b \otimes e_b \,:\, b \in \Sigma\} \tag{4.134}$$

for every $a \in \Gamma$. This is equivalent to the condition that A_a is diagonal for every $a \in \Gamma$, and so it has been proved that statement 1 implies statement 3.

Statement 3 trivially implies statement 2, so it remains to prove that statement 2 implies statement 1. For a Kraus representation of Φ having the form (4.129), where Γ is an alphabet and $\{A_a : a \in \Gamma\}$ is a collection of diagonal operators, let $\{v_a : a \in \Gamma\} \subset \mathcal{X}$ be the collection of vectors satisfying $A_a = \mathrm{Diag}(v_a)$ for each $a \in \Gamma$, and define

$$P = \sum_{a \in \Gamma} v_a v_a^*. \tag{4.135}$$

A calculation reveals that

$$P \odot X = \sum_{a \in \Gamma} \sum_{b,c \in \Sigma} X(b,c)\, v_a(b) \overline{v_a(c)}\, E_{b,c} = \sum_{a \in \Gamma} A_a X A_a^* \tag{4.136}$$

for every $X \in \mathrm{L}(\mathcal{X})$. It has therefore been proved that Φ is a Schur mapping, so statement 2 implies statement 1 as required. □

4.2 General properties of unital channels

This section proves a few basic facts holding for unital channels in general. In particular, the extreme points of the set of all unital channels defined with respect to a given space are characterized, and properties relating to fixed-points and norms of unital channels are established.

4.2.1 Extreme points of the set of unital channels

Theorem 2.31 provides a criterion through which one may determine if a given channel $\Phi \in \mathrm{C}(\mathcal{X})$ is an extreme point of the set of all channels $\mathrm{C}(\mathcal{X})$, based on any linearly independent set of Kraus operators of Φ. It will be demonstrated by Theorem 4.21 below that a similar criterion holds when the set $\mathrm{C}(\mathcal{X})$ is replaced by the set of all unital channels

$$\big\{\Phi \in \mathrm{C}(\mathcal{X}) \,:\, \Phi(\mathbb{1}_{\mathcal{X}}) = \mathbb{1}_{\mathcal{X}}\big\}. \tag{4.137}$$

Indeed, the analogous theorem for unital channels will follow directly from Theorem 2.31, together with an embedding of the set (4.137) within the set of all channels of the form $\mathrm{C}(\mathcal{X} \oplus \mathcal{X})$.

Assume that a complex Euclidean space $\mathcal{X}$ has been fixed, and define an operator

$$V \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X}, (\mathcal{X} \oplus \mathcal{X}) \otimes (\mathcal{X} \oplus \mathcal{X})) \tag{4.138}$$

by the equation

$$V \operatorname{vec}(X) = \operatorname{vec}\begin{pmatrix} X & 0 \\ 0 & X^{\mathsf{T}} \end{pmatrix} \tag{4.139}$$

holding for all operators $X \in \mathrm{L}(\mathcal{X})$. It may be verified that $V^*V = 2\mathbb{1}_{\mathcal{X}}$. For every map $\Phi \in \mathrm{T}(\mathcal{X})$, define $\phi(\Phi) \in \mathrm{T}(\mathcal{X} \oplus \mathcal{X})$ to be the unique map for which the equation

$$J(\phi(\Phi)) = VJ(\Phi)V^* \tag{4.140}$$

holds, and observe that the mapping $\phi : \mathrm{T}(\mathcal{X}) \to \mathrm{T}(\mathcal{X} \oplus \mathcal{X})$ defined in this way is injective and linear. If $\Phi \in \mathrm{T}(\mathcal{X})$ is defined by a Kraus representation

$$\Phi(X) = \sum_{a \in \Sigma} A_a X B_a^*, \tag{4.141}$$

then it holds that

$$\phi(\Phi)\begin{pmatrix} X_{0,0} & X_{0,1} \\ X_{1,0} & X_{1,1} \end{pmatrix} = \sum_{a \in \Sigma} \begin{pmatrix} A_a & 0 \\ 0 & A_a^{\mathsf{T}} \end{pmatrix} \begin{pmatrix} X_{0,0} & X_{0,1} \\ X_{1,0} & X_{1,1} \end{pmatrix} \begin{pmatrix} B_a & 0 \\ 0 & B_a^{\mathsf{T}} \end{pmatrix}^* \tag{4.142}$$

is a Kraus representation of $\phi(\Phi)$. The following observations concerning the mapping $\phi : \mathrm{T}(\mathcal{X}) \to \mathrm{T}(\mathcal{X} \oplus \mathcal{X})$ may be verified:

1. A map $\Phi \in \mathrm{T}(\mathcal{X})$ is completely positive if and only if $\phi(\Phi) \in \mathrm{T}(\mathcal{X} \oplus \mathcal{X})$ is completely positive.
2. A map $\Phi \in \mathrm{T}(\mathcal{X})$ is both trace-preserving and unital if and only if $\phi(\Phi) \in \mathrm{T}(\mathcal{X} \oplus \mathcal{X})$ is trace-preserving.

In particular, $\Phi \in \mathrm{C}(\mathcal{X})$ is a unital channel if and only if $\phi(\Phi) \in \mathrm{C}(\mathcal{X} \oplus \mathcal{X})$ is a channel. (In this case, $\phi(\Phi)$ will also happen to be unital.)

Lemma 4.20. *Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{C}(\mathcal{X})$ be a unital channel, and let $\phi(\Phi) \in \mathrm{C}(\mathcal{X} \oplus \mathcal{X})$ be the channel defined from Φ by the equation (4.140). It holds that Φ is an extreme point in the set of all unital channels in $\mathrm{C}(\mathcal{X})$ if and only if $\phi(\Phi)$ is an extreme point of the set of channels $\mathrm{C}(\mathcal{X} \oplus \mathcal{X})$.*

Proof. Suppose first that Φ is not an extreme point in the set of all unital channels in $\mathrm{C}(\mathcal{X})$, so that

$$\Phi = \lambda \Psi_0 + (1 - \lambda)\Psi_1 \tag{4.143}$$

for distinct unital channels $\Psi_0, \Psi_1 \in \mathrm{C}(\mathcal{X})$ and a scalar $\lambda \in (0, 1)$. As the mapping ϕ is linear and injective, it therefore holds that

$$\phi(\Phi) = \lambda \phi(\Psi_0) + (1 - \lambda)\phi(\Psi_1), \tag{4.144}$$

which is a proper convex combination of distinct channels. This implies that $\phi(\Phi)$ is not an extreme point of the set of channels $\mathrm{C}(\mathcal{X} \oplus \mathcal{X})$.

Suppose, on the other hand, that $\phi(\Phi)$ is not an extreme point of the set of channels $\mathrm{C}(\mathcal{X} \oplus \mathcal{X})$, so that

$$\phi(\Phi) = \lambda \Xi_0 + (1 - \lambda)\Xi_1 \tag{4.145}$$

for distinct channels $\Xi_0, \Xi_1 \in \mathrm{C}(\mathcal{X} \oplus \mathcal{X})$ and a scalar $\lambda \in (0, 1)$. Taking the Choi representations of both sides of this equation yields

$$VJ(\Phi)V^* = \lambda J(\Xi_0) + (1 - \lambda)J(\Xi_1). \tag{4.146}$$

It therefore follows from Lemma 2.30 that

$$J(\Xi_0) = VQ_0V^* \quad \text{and} \quad J(\Xi_1) = VQ_1V^* \tag{4.147}$$

for some choice of positive semidefinite operators $Q_0, Q_1 \in \mathrm{Pos}(\mathcal{X})$. Letting $\Psi_0, \Psi_1 \in \mathrm{T}(\mathcal{X})$ be the maps defined by $J(\Psi_0) = Q_0$ and $J(\Psi_1) = Q_1$, one has $\Xi_0 = \phi(\Psi_0)$ and $\Xi_1 = \phi(\Psi_1)$. As $\phi(\Psi_0) = \Xi_0$ and $\phi(\Psi_1) = \Xi_1$ are distinct channels, it follows that Ψ_0 and Ψ_1 are distinct unital channels. It holds that

$$\phi(\Phi) = \lambda\phi(\Psi_0) + (1-\lambda)\phi(\Psi_1) \tag{4.148}$$

and therefore

$$\Phi = \lambda\Psi_0 + (1-\lambda)\Psi_1, \tag{4.149}$$

which implies that Φ is not an extreme point in the set of all unital channels in $\mathrm{C}(\mathcal{X})$. □

Theorem 4.21. *Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{C}(\mathcal{X})$ be a unital channel, and let $\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$ be a linearly independent set of operators satisfying*

$$\Phi(X) = \sum_{a\in\Sigma} A_a X A_a^* \tag{4.150}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The channel Φ is an extreme point in the set of all unital channels in $\mathrm{C}(\mathcal{X})$ if and only if the collection

$$\left\{ \begin{pmatrix} A_b^* A_a & 0 \\ 0 & A_a A_b^* \end{pmatrix} : (a,b) \in \Sigma \times \Sigma \right\} \tag{4.151}$$

of operators is linearly independent.

Proof. By Lemma 4.20, the channel Φ is an extreme point of the set of unital channels in $\mathrm{C}(\mathcal{X})$ if and only if the channel $\phi(\Phi)$ is an extreme point in the set $\mathrm{C}(\mathcal{X} \oplus \mathcal{X})$, for $\phi : \mathrm{T}(\mathcal{X}) \to \mathrm{T}(\mathcal{X} \oplus \mathcal{X})$ being the mapping defined by the equation (4.140). By Theorem 2.31, it follows that $\phi(\Phi)$ is an extreme point of the set of channels $\mathrm{C}(\mathcal{X} \oplus \mathcal{X})$ if and only if

$$\left\{ \begin{pmatrix} A_b^* A_a & 0 \\ 0 & \overline{A_b} A_a^{\mathsf{T}} \end{pmatrix} : (a,b) \in \Sigma \times \Sigma \right\} \tag{4.152}$$

is a linearly independent collection of operators. Taking the transpose of the lower-right-hand block, which does not change whether or not the set is linearly independent, it follows that $\phi(\Phi)$ is an extreme point of the set $\mathrm{C}(\mathcal{X} \oplus \mathcal{X})$ if and only if the set (4.151) is linearly independent. □

Unital qubit channels are mixed-unitary

There exist non-mixed-unitary unital channels, as shown in Example 4.3. The existence of such channels, however, requires that the underlying space has dimension at least 3; when Theorem 4.21 is combined with the following lemma, one concludes that every unital qubit channel is mixed-unitary.

Lemma 4.22. *Let $\mathcal{X}$ be a complex Euclidean space and let $A_0, A_1 \in \mathrm{L}(\mathcal{X})$ be operators such that*

$$A_0^* A_0 + A_1^* A_1 = \mathbb{1}_{\mathcal{X}} = A_0 A_0^* + A_1 A_1^*. \tag{4.153}$$

There exist unitary operators $U, V \in \mathrm{U}(\mathcal{X})$ such that VA_0U^* and VA_1U^* are diagonal operators.

Proof. It suffices to prove that there exists a unitary operator $W \in \mathrm{U}(\mathcal{X})$ such that the operators WA_0 and WA_1 are both normal and satisfy

$$[WA_0, WA_1] = 0, \tag{4.154}$$

for then it follows by Theorem 1.5 that one may choose U so that UWA_0U^* and UWA_1U^* are diagonal, then take $V = UW$.

Let $U_0, U_1 \in \mathrm{U}(\mathcal{X})$ and $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ be operators providing the polar decompositions

$$A_0 = U_0 P_0 \quad \text{and} \quad A_1 = U_1 P_1, \tag{4.155}$$

and let $W = U_0^*$. It holds that $WA_0 = P_0$, which is positive semidefinite and therefore normal. To verify that WA_1 is normal, observe that the assumption (4.153) implies

$$U_1 P_1^2 U_1^* = \mathbb{1} - U_0 P_0^2 U_0^* \quad \text{and} \quad P_1^2 = \mathbb{1} - P_0^2, \tag{4.156}$$

and therefore

$$\begin{aligned} (WA_1)(WA_1)^* &= U_0^* U_1 P_1^2 U_1^* U_0 = U_0^* (\mathbb{1} - U_0 P_0^2 U_0^*) U_0 \\ &= \mathbb{1} - P_0^2 = P_1^2 = P_1 U_1^* U_0 U_0^* U_1 P_1 = (WA_1)^* (WA_1). \end{aligned} \tag{4.157}$$

It remains to prove that the operators WA_0 and WA_1 commute. It follows from the equation $P_1^2 = \mathbb{1} - P_0^2$ that P_0^2 and P_1^2 commute. As P_0^2 and P_1^2 are commuting positive semidefinite operators, it therefore holds that P_0 and P_1 commute. Substituting $P_1^2 = \mathbb{1} - P_0^2$ into the equation

$$U_1 P_1^2 U_1^* = \mathbb{1} - U_0 P_0^2 U_0^*, \tag{4.158}$$

one finds that

$$U_0 P_0^2 U_0^* = U_1 P_0^2 U_1^*, \tag{4.159}$$

and therefore, by taking the square root of both sides of this equation,

$$U_0 P_0 U_0^* = U_1 P_0 U_1^*. \tag{4.160}$$

This implies that

$$P_0 U_0^* U_1 = U_0^* U_1 P_0, \tag{4.161}$$

and therefore P_0 and $U_0^* U_1$ commute. It follows that

$$\begin{aligned} (WA_0)(WA_1) &= P_0 U_0^* U_1 P_1 = U_0^* U_1 P_0 P_1 \\ &= U_0^* U_1 P_1 P_0 = (WA_1)(WA_0), \end{aligned} \tag{4.162}$$

and so WA_0 and WA_1 commute as required. □

Theorem 4.23. *Let $\mathcal{X}$ be a complex Euclidean space with $\dim(\mathcal{X}) = 2$. Every unital channel $\Phi \in \mathrm{C}(\mathcal{X})$ is a mixed-unitary channel.*

Proof. The set

$$\{\Phi \in \mathrm{C}(\mathcal{X}) : \Phi(\mathbb{1}_{\mathcal{X}}) = \mathbb{1}_{\mathcal{X}}\} \tag{4.163}$$

of unital channels, defined with respect to the space $\mathcal{X}$, is both compact and convex—both of these properties are consequences of the fact that this set is equal to the intersection of the compact and convex set $\mathrm{C}(\mathcal{X})$ with the (closed) affine subspace of all maps $\Phi \in \mathrm{T}(\mathcal{X})$ satisfying $\Phi(\mathbb{1}_{\mathcal{X}}) = \mathbb{1}_{\mathcal{X}}$. As this set is compact and convex, Theorem 1.10 implies that it is equal to the convex hull of its extreme points. To complete the proof, it therefore suffices to establish that every unital channel $\Phi \in \mathrm{C}(\mathcal{X})$ that is not a unitary channel is not an extreme point of the set (4.163).

Toward this goal, let $\Phi \in \mathrm{C}(\mathcal{X})$ be an arbitrary unital channel, and let $\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$ be a linearly independent collection of operators satisfying

$$\Phi(X) = \sum_{a \in \Sigma} A_a X A_a^* \tag{4.164}$$

for all $X \in \mathrm{L}(\mathcal{X})$. One has that Φ is a unitary channel if and only if $|\Sigma| = 1$, so it suffices to prove that Φ is not an extreme point of the set (4.163) whenever $|\Sigma| \geq 2$.

By Theorem 4.21, the channel Φ is an extreme point of the set (4.163) if and only if

$$\left\{ \begin{pmatrix} A_b^* A_a & 0 \\ 0 & A_a A_b^* \end{pmatrix} \,:\, (a,b) \in \Sigma \times \Sigma \right\} \subset \mathrm{L}(\mathcal{X} \oplus \mathcal{X}) \tag{4.165}$$

is a linearly independent collection of operators. There are two cases that must be considered: the first case is that $|\Sigma| \geq 3$ and the second case is that $|\Sigma| = 2$.

For the first case, one has that the collection (4.165) includes at least 9 operators drawn from the 8-dimensional subspace

$$\left\{ \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix} \,:\, X, Y \in \mathrm{L}(\mathcal{X}) \right\}. \tag{4.166}$$

Thus, if $|\Sigma| \geq 3$, then the collection (4.165) cannot be linearly independent, and therefore Φ is not an extreme point of the set (4.163).

It remains to consider the case $|\Sigma| = 2$. There is no loss of generality in assuming $\Sigma = \{0,1\}$ and $\mathcal{X} = \mathbb{C}^{\Sigma}$. By Lemma 4.22, there must exist unitary operators $U, V \in \mathrm{U}(\mathcal{X})$ such that VA_0U^* and VA_1U^* are diagonal operators:

$$\begin{aligned} VA_0U^* &= \alpha_0 E_{0,0} + \beta_0 E_{1,1}\,, \\ VA_1U^* &= \alpha_1 E_{0,0} + \beta_1 E_{1,1}. \end{aligned} \tag{4.167}$$

The following equations therefore hold for every choice of $a, b \in \Sigma$:

$$\begin{aligned} A_b^* A_a &= \alpha_a \overline{\alpha_b} U^* E_{0,0} U + \beta_a \overline{\beta_b} U^* E_{1,1} U, \\ A_a A_b^* &= \alpha_a \overline{\alpha_b} V^* E_{0,0} V + \beta_a \overline{\beta_b} V^* E_{1,1} V. \end{aligned} \tag{4.168}$$

The set (4.165) is therefore contained in the subspace spanned by the set of operators

$$\left\{ \begin{pmatrix} U^* E_{0,0} U & 0 \\ 0 & V^* E_{0,0} V \end{pmatrix}, \begin{pmatrix} U^* E_{1,1} U & 0 \\ 0 & V^* E_{1,1} V \end{pmatrix} \right\}. \tag{4.169}$$

The collection (4.165) contains 4 operators drawn from a two-dimensional space, and therefore cannot be linearly independent. This implies that the channel Φ is not an extreme point of the set (4.163), which completes the proof. □

4.2.2 Fixed-points, spectra, and norms of unital channels

Every channel of the form $\Phi \in \mathrm{C}(\mathcal{X})$ must have at least one density operator fixed point, meaning a density operator $\rho \in \mathrm{D}(\mathcal{X})$ satisfying

$$\Phi(\rho) = \rho. \tag{4.170}$$

One may see this fact as a consequence of the Brouwer fixed-point theorem, which states that every continuous function mapping a compact, convex set in a Euclidean space to itself must have a fixed point. The full power of the Brouwer fixed-point theorem is, however, really not needed in this case; the fact that channels are linear maps allows for a simpler proof. The following theorem establishes this fact in slightly greater generality, for any positive and trace-preserving map $\Phi \in \mathrm{T}(\mathcal{X})$.

Theorem 4.24. *Let $\mathcal{X}$ be a complex Euclidean space and let $\Phi \in \mathrm{T}(\mathcal{X})$ be a positive and trace-preserving map. There exists a density operator $\rho \in \mathrm{D}(\mathcal{X})$ such that $\Phi(\rho) = \rho$.*

Proof. For each nonnegative integer $n \in \mathbb{N}$, define a map $\Psi_n \in \mathrm{T}(\mathcal{X})$ as

$$\Psi_n(X) = \frac{1}{2^n} \sum_{k=0}^{2^n - 1} \Phi^k(X) \tag{4.171}$$

for each $X \in \mathrm{L}(\mathcal{X})$, and define a set

$$\mathcal{C}_n = \{\Psi_n(\rho) \,:\, \rho \in \mathrm{D}(\mathcal{X})\}. \tag{4.172}$$

As Φ is linear, positive, and trace-preserving, the same is true of Ψ_n, and so it follows that $\mathcal{C}_n$ is a compact and convex subset of $\mathrm{D}(\mathcal{X})$ for each $n \in \mathbb{N}$. By the convexity of the set $\mathcal{C}_n$, it holds that

$$\Psi_{n+1}(\rho) = \frac{1}{2}\Psi_n(\rho) + \frac{1}{2}\Psi_n\Big(\Phi^{2^n}(\rho)\Big) \in \mathcal{C}_n \tag{4.173}$$

for every $\rho \in \mathrm{D}(\mathcal{X})$, and therefore $\mathcal{C}_{n+1} \subseteq \mathcal{C}_n$, for every $n \in \mathbb{N}$. As each $\mathcal{C}_n$ is compact and $\mathcal{C}_{n+1} \subseteq \mathcal{C}_n$ for all $n \in \mathbb{N}$, it follows that there must exist an element

$$\rho \in \bigcap_{n \in \mathbb{N}} \mathcal{C}_n \tag{4.174}$$

contained in the intersection of all of these sets.

Now, fix any choice of ρ satisfying (4.174). For an arbitrary choice of $n \in \mathbb{N}$, it holds that $\rho = \Psi_n(\sigma)$ for some choice of $\sigma \in \mathrm{D}(\mathcal{X})$, and therefore

$$\Phi(\rho) - \rho = \Phi(\Psi_n(\sigma)) - \Psi_n(\sigma) = \frac{\Phi^{2^n}(\sigma) - \sigma}{2^n}. \tag{4.175}$$

As the trace distance between two density operators cannot exceed 2, it follows that

$$\bigl\| \Phi(\rho) - \rho \bigr\|_1 \leq \frac{1}{2^{n-1}}. \tag{4.176}$$

This bound holds for every $n \in \mathbb{N}$, which implies $\| \Phi(\rho) - \rho \|_1 = 0$, and therefore $\Phi(\rho) = \rho$ as required. □

There is, of course, no difficulty in proving the existence of a density operator fixed point of a unital channel: if $\Phi \in \mathrm{C}(\mathcal{X})$ is a unital channel, then $\omega = \mathbb{1}_{\mathcal{X}} / \dim(\mathcal{X})$ is a density operator fixed point of Φ. What is more interesting is the fact that the collection of all operators $X \in \mathrm{L}(\mathcal{X})$ satisfying $\Phi(X) = X$ forms a unital sub-algebra of $\mathrm{L}(\mathcal{X})$, as the following theorem implies.

Theorem 4.25. *Let $\mathcal{X}$ be a complex Euclidean space and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a unital channel. Suppose further that Σ is an alphabet and $\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$ is a collection of operators yielding a Kraus representation of Φ:*

$$\Phi(X) = \sum_{a \in \Sigma} A_a X A_a^* \tag{4.177}$$

for all $X \in \mathrm{L}(\mathcal{X})$. For every $X \in \mathrm{L}(\mathcal{X})$ it holds that $\Phi(X) = X$ if and only if $[X, A_a] = 0$ for every $a \in \Sigma$.

Proof. If $X \in \mathrm{L}(\mathcal{X})$ is an operator for which $[X, A_a] = 0$ for every $a \in \Sigma$, then

$$\Phi(X) = \sum_{a \in \Sigma} A_a X A_a^* = \sum_{a \in \Sigma} X A_a A_a^* = X \Phi(\mathbb{1}) = X, \tag{4.178}$$

where the last equality follows from the assumption that Φ is unital.

Now suppose that $X \in \mathrm{L}(\mathcal{X})$ is an operator for which $\Phi(X) = X$, and consider the positive semidefinite operator

$$\sum_{a \in \Sigma} [X, A_a] \, [X, A_a]^*. \tag{4.179}$$

Expanding this operator and using the assumptions that Φ is unital and $\Phi(X) = X$ (and therefore $\Phi(X^*) = X^*$), one has

$$\begin{aligned}
&\sum_{a\in\Sigma} [X, A_a]\,[X, A_a]^* \\
&\quad = \sum_{a\in\Sigma} \Big(\big(XA_a - A_aX\big)\big(A_a^*X^* - X^*A_a^*\big) \Big) \\
&\quad = \sum_{a\in\Sigma} \big(XA_aA_a^*X^* - A_aXA_a^*X^* - XA_aX^*A_a^* + A_aXX^*A_a^*\big) \\
&\quad = XX^* - \Phi(X)X^* - X\Phi(X^*) + \Phi(XX^*) \\
&\quad = \Phi(XX^*) - XX^*.
\end{aligned} \tag{4.180}$$

As Φ is a channel, and is therefore trace-preserving, it holds that the trace of the operator represented by the previous equation is zero. The only traceless positive semidefinite operator is the zero operator, and therefore

$$\sum_{a\in\Sigma} [X, A_a]\,[X, A_a]^* = 0. \tag{4.181}$$

This implies that each of the terms $[X, A_a]\,[X, A_a]^*$ is zero, and therefore each operator $[X, A_a]$ is zero. □

For any channel of the form $\Phi \in \mathrm{C}(\mathcal{X})$, for $\mathcal{X}$ being a complex Euclidean space, one has that the natural representation of Φ is a square operator of the form $K(\Phi) \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$. The following proposition establishes that the spectral radius of $K(\Phi)$ is necessarily equal to 1.

Proposition 4.26. *Let $\mathcal{X}$ be a complex Euclidean space and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a channel. The spectral radius of $K(\Phi)$ is equal to 1.*

Proof. By Theorem 4.24, there must exist a density operator $\rho \in \mathrm{D}(\mathcal{X})$ such that $\Phi(\rho) = \rho$, which implies that $K(\Phi)$ has an eigenvalue equal to 1.

It remains to prove that every eigenvalue of $K(\Phi)$ is at most 1 in absolute value, which is equivalent to the statement that $|\lambda| \leq 1$ for every choice of a nonzero operator $X \in \mathrm{L}(\mathcal{X})$ and a complex number $\lambda \in \mathbb{C}$ satisfying

$$\Phi(X) = \lambda X. \tag{4.182}$$

Suppose that $X \in \mathrm{L}(\mathcal{X})$ and $\lambda \in \mathbb{C}$ satisfy (4.182). By Corollary 3.43, it holds that $\|\Phi\|_1 = 1$, and therefore

$$1 \geq \frac{\|\Phi(X)\|_1}{\|X\|_1} = \frac{\|\lambda X\|_1}{\|X\|_1} = |\lambda|. \tag{4.183}$$

The required bound on λ holds, which completes the proof. □

While the spectral radius of the natural representation $K(\Phi)$ of every channel $\Phi \in \mathrm{C}(\mathcal{X})$ must equal 1, it will not generally be the case that the spectral norm of $K(\Phi)$ will be 1. As the following theorem establishes, this happens if and only if Φ is a unital channel. Like Theorem 4.24, the property of complete positivity is not needed in the proof of this fact, and so it holds not only for channels, but for all positive and trace-preserving maps.

Theorem 4.27. *Let $\mathcal{X}$ be a complex Euclidean space and let $\Phi \in \mathrm{T}(\mathcal{X})$ be a positive and trace-preserving map. It holds that Φ is unital if and only if $\|K(\Phi)\| = 1$.*

Proof. Suppose first that Φ is a unital channel. It is evident that $\|K(\Phi)\| \geq 1$, as Theorem 4.24 implies that

$$K(\Phi)\operatorname{vec}(\rho) = \operatorname{vec}(\rho) \tag{4.184}$$

for some choice of a density operator $\rho \in \mathrm{D}(\mathcal{X})$. It therefore suffices to prove that $\|K(\Phi)\| \leq 1$, which is equivalent to the condition that

$$\big\|\Phi(X)\big\|_2 \leq \big\|X\big\|_2 \tag{4.185}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

Consider first an arbitrary Hermitian operator $H \in \mathrm{Herm}(\mathcal{X})$. Let

$$H = \sum_{k=1}^{n} \lambda_k x_k x_k^* \tag{4.186}$$

be a spectral decomposition of H, for $n = \dim(\mathcal{X})$, and let

$$\rho_k = \Phi(x_k x_k^*) \tag{4.187}$$

for each $k \in \{1, \ldots, n\}$. One has that $\rho_1, \ldots, \rho_n$ are density operators, as a consequence of the fact that Φ is positive and trace-preserving. Moreover, as Φ is unital, it follows that $\rho_1 + \cdots + \rho_n = \mathbb{1}$. It holds that

$$\big\|\Phi(H)\big\|_2^2 = \big\|\lambda_1\rho_1 + \cdots + \lambda_n\rho_n\big\|_2^2 = \sum_{1\leq j,k\leq n} \lambda_j\lambda_k\langle\rho_j,\rho_k\rangle. \tag{4.188}$$

The Cauchy–Schwarz inequality implies that

$$\begin{aligned} &\sum_{1\leq j,k\leq n} \lambda_j\lambda_k\langle\rho_j,\rho_k\rangle \\ &\leq \sqrt{\sum_{1\leq j,k\leq n} \lambda_j^2\langle\rho_j,\rho_k\rangle}\sqrt{\sum_{1\leq j,k\leq n} \lambda_k^2\langle\rho_j,\rho_k\rangle} = \sum_{k=1}^{n}\lambda_k^2 = \big\|H\big\|_2^2, \end{aligned} \tag{4.189}$$

where the first equality has followed from the fact that $\rho_1 + \cdots + \rho_n = \mathbb{1}$. It has therefore been established that $\|\Phi(H)\|_2 \leq \|H\|_2$ for all Hermitian operators $H \in \mathrm{Herm}(\mathcal{X})$.

Now consider any operator $X \in \mathrm{L}(\mathcal{X})$, written as $X = H + iK$ for

$$H = \frac{X + X^*}{2} \quad \text{and} \quad K = \frac{X - X^*}{2i} \tag{4.190}$$

being Hermitian operators, and observe that

$$\|X\|_2^2 = \|H\|_2^2 + \|K\|_2^2. \tag{4.191}$$

As Φ is necessarily Hermiticity-preserving, one finds that

$$\|\Phi(X)\|_2^2 = \|\Phi(H) + i\Phi(K)\|_2^2 = \|\Phi(H)\|_2^2 + \|\Phi(K)\|_2^2. \tag{4.192}$$

Therefore

$$\|\Phi(X)\|_2^2 = \|\Phi(H)\|_2^2 + \|\Phi(K)\|_2^2 \leq \|H\|_2^2 + \|K\|_2^2 = \|X\|_2^2, \tag{4.193}$$

so $\|\Phi(X)\|_2 \leq \|X\|_2$, as required.

Now suppose that Φ is a positive and trace-preserving map for which $\|K(\Phi)\| = 1$, which is equivalent to the condition that $\|\Phi(X)\|_2 \leq \|X\|_2$ for every $X \in \mathrm{L}(\mathcal{X})$. In particular, it must hold that

$$\|\Phi(\mathbb{1})\|_2 \leq \|\mathbb{1}\|_2 = \sqrt{n}, \tag{4.194}$$

for $n = \dim(\mathcal{X})$. As Φ is positive and trace-preserving, one has that $\Phi(\mathbb{1})$ is positive semidefinite and has trace equal to n. When these observations are combined with the Cauchy–Schwarz inequality, one finds that

$$n = \mathrm{Tr}\big(\Phi(\mathbb{1})\big) = \langle \mathbb{1}, \Phi(\mathbb{1}) \rangle \leq \|\mathbb{1}\|_2 \, \|\Phi(\mathbb{1})\|_2 \leq n. \tag{4.195}$$

Equality is therefore obtained in the Cauchy–Schwarz inequality, implying that $\Phi(\mathbb{1})$ and $\mathbb{1}$ are linearly dependent. As $\mathrm{Tr}(\mathbb{1}) = \mathrm{Tr}(\Phi(\mathbb{1}))$, it follows that $\Phi(\mathbb{1})$ and $\mathbb{1}$ must in fact be equal, and therefore Φ is unital. □

4.3 Majorization

This section introduces the *majorization* relation for Hermitian operators, which is a generalization of a similar concept for real vectors. Intuitively speaking, the majorization relation formalizes the notion of one object being obtained from another through a "random mixing process."

One may formalize the majorization relation, both for real vectors and for Hermitian operators, in multiple, equivalent ways. Once formalized, it is a very useful mathematical concept. In the theory of quantum information, majorization has a particularly striking application in the form of Nielsen's theorem (Theorem 6.37 in Chapter 6), which gives a precise characterization of the possible transformations between bipartite pure states that may be performed by two individuals whose communications with one another are restricted to classical information transmissions.

4.3.1 Majorization for real vectors

The definition of the majorization relation for real vectors to be presented in this book is based on the class of *doubly stochastic* operators. A discussion of such operators follows, after which the majorization relation for real vectors is defined.

Doubly stochastic operators

Let Σ be an alphabet, and consider the real Euclidean space $\mathbb{R}^{\Sigma}$. An operator $A \in \mathrm{L}(\mathbb{R}^{\Sigma})$ acting on this vector space is said to be *stochastic* if and only if

1. $A(a,b) \geq 0$ for each $(a,b) \in \Sigma \times \Sigma$, and
2. $\sum_{a \in \Sigma} A(a,b) = 1$ for each $b \in \Sigma$.

This condition is equivalent to Ae_b being a probability vector for each $b \in \Sigma$, or equivalently, that A maps probability vectors to probability vectors. An operator $A \in \mathrm{L}(\mathbb{R}^{\Sigma})$ is said to be *doubly stochastic* if and only if

1. $A(a,b) \geq 0$ for each $(a,b) \in \Sigma \times \Sigma$,
2. $\sum_{a \in \Sigma} A(a,b) = 1$ for each $b \in \Sigma$, and
3. $\sum_{b \in \Sigma} A(a,b) = 1$ for each $a \in \Sigma$.

That is, an operator A is doubly stochastic if and only if both A and A^{T} (or, equivalently, both A and A^*) are stochastic, which is equivalent to the condition that every row and every column of the matrix representation of A forms a probability vector.

Doubly stochastic operators have a close relationship to *permutation operators*. For each permutation $\pi \in \mathrm{Sym}(\Sigma)$, one defines the permutation operator $V_\pi \in \mathrm{L}(\mathbb{R}^\Sigma)$ as

$$V_\pi(a,b) = \begin{cases} 1 & \text{if } a = \pi(b) \\ 0 & \text{otherwise} \end{cases} \tag{4.196}$$

for every $(a,b) \in \Sigma \times \Sigma$. Equivalently, V_π is the unique operator satisfying the equation $V_\pi e_b = e_{\pi(b)}$ for each $b \in \Sigma$. It is evident that permutation operators are doubly stochastic. The next theorem establishes that the set of all doubly stochastic operators is, in fact, equal to the convex hull of the permutation operators.

Theorem 4.28 (Birkhoff–von Neumann theorem)**.** *Let Σ be an alphabet and let $A \in \mathrm{L}(\mathbb{R}^\Sigma)$ be an operator. It holds that A is doubly stochastic if and only if there exists a probability vector $p \in \mathcal{P}(\mathrm{Sym}(\Sigma))$ such that*

$$A = \sum_{\pi \in \mathrm{Sym}(\Sigma)} p(\pi) V_\pi. \tag{4.197}$$

Proof. The set of all doubly stochastic operators acting on $\mathbb{R}^\Sigma$ is convex and compact, and is therefore equal to the convex hull of its extreme points by Theorem 1.10. The theorem will therefore follow from the demonstration that every extreme point in this set is a permutation operator. With this fact in mind, let A be a doubly stochastic operator that is not a permutation operator. It will be proved that A is not an extreme point of the set of doubly stochastic operators, which is sufficient to complete the proof.

Given that A is doubly stochastic but not a permutation operator, there must exist at least one pair $(a_1, b_1) \in \Sigma \times \Sigma$ such that $A(a_1, b_1) \in (0,1)$. As $\sum_b A(a_1, b) = 1$ and $A(a_1, b_1) \in (0,1)$, one may conclude that there exists an index $b_2 \neq b_1$ such that $A(a_1, b_2) \in (0,1)$. Applying similar reasoning, but to the first index rather than the second, it follows that there must exist an index $a_2 \neq a_1$ such that $A(a_2, b_2) \in (0,1)$. Repeating this argument, one may eventually find a closed loop of even length among the entries of A that are contained in the interval $(0,1)$, alternating between the first and

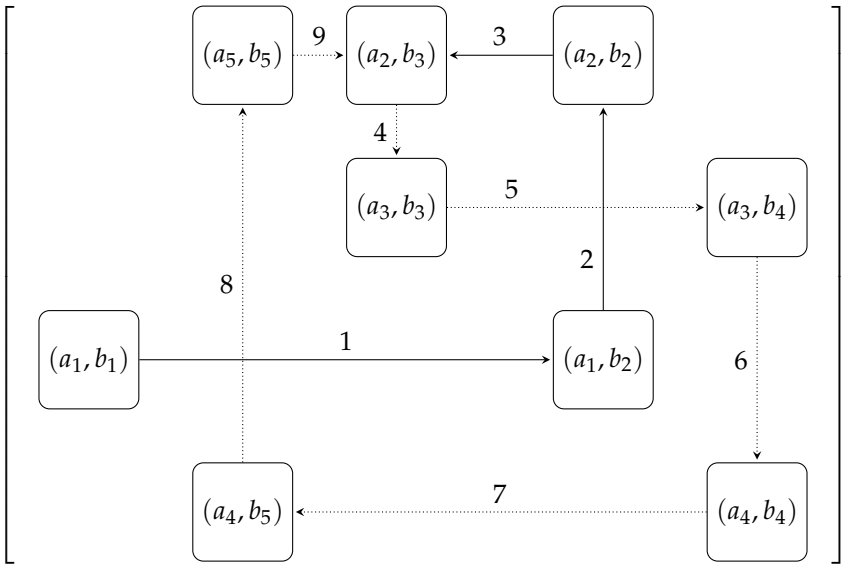


Figure 4.1: An example of a closed loop consisting of entries of A that are contained in the interval $(0, 1)$. The loop is indicated by the dotted lines.

second indices (i.e., between rows and columns). A loop must eventually be formed, given that there are only finitely many entries in the matrix A; and an odd-length loop can be avoided by an appropriate choice for the entry that closes the loop. This process is illustrated in Figure 4.1.

Let $\varepsilon \in (0, 1)$ be equal to the minimum value over the entries in a closed loop of the form just described, and define B to be the operator obtained by setting each entry in the closed loop to be $\pm\varepsilon$, alternating sign among the entries as suggested in Figure 4.2. All of the other entries in B are set to 0. Finally, consider the operators $A + B$ and $A - B$. As A is doubly stochastic and the row and column sums of B are all 0, it holds that both $A + B$ and $A - B$ also have row and column sums equal to 1. As ε was chosen to be no larger than the smallest entry within the chosen closed loop, none of the entries of $A + B$ or $A - B$ are negative, and therefore $A - B$ and $A + B$ are doubly stochastic. As B is nonzero, it holds that $A + B$ and $A - B$ are distinct. Thus,

$$A = \frac{1}{2}(A + B) + \frac{1}{2}(A - B) \tag{4.198}$$

is a proper convex combination of doubly stochastic operators, and is therefore not an extreme point in the set of doubly stochastic operators. □

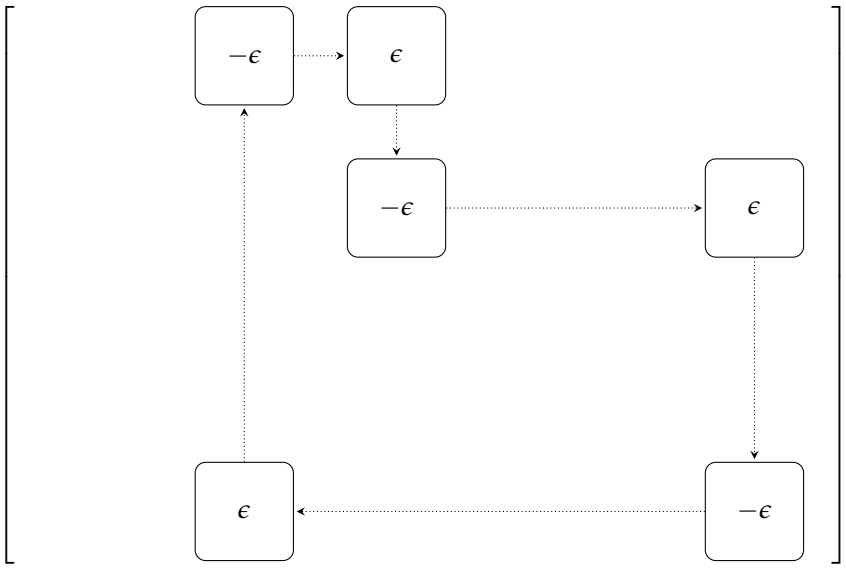


Figure 4.2: The operator B. All entries besides those indicated are 0.

Definition and characterizations of majorization for real vectors

A definition of the majorization relation for vectors of real numbers, based on the actions of doubly stochastic operators, is as follows.

Definition 4.29. Let Σ be an alphabet and let $u, v \in \mathbb{R}^{\Sigma}$ be vectors. It is said that u *majorizes* v, written $v \prec u$, if and only if there exists a doubly stochastic operator $A \in \mathrm{L}\big(\mathbb{R}^{\Sigma}\big)$ for which $v = Au$.

By the Birkhoff–von Neumann theorem (Theorem 4.28), one may view this definition as formalizing the sort of "random mixing process" suggested at the beginning of the current section. An operator A is doubly stochastic if and only if it is equal to a convex combination of permutation operators, so the relation $v \prec u$ holds precisely when v can be obtained by randomly choosing a permutation $\pi \in \mathrm{Sym}(\Sigma)$, with respect to a chosen distribution $p \in \mathcal{P}(\mathrm{Sym}(\Sigma))$, shuffling the entries of u in accordance with the chosen permutation π, and then averaging the resulting vectors with respect to p.

The following theorem provides two alternative characterizations of the majorization relation for real vectors. The statement of the theorem makes use of the following notation: for every vector $u \in \mathbb{R}^{\Sigma}$ and for $n = |\Sigma|$, one

writes

$$r(u) = (r_1(u), \ldots, r_n(u)) \tag{4.199}$$

to denote the vector obtained by *sorting* the entries of u in decreasing order. In other words, one has

$$\{u(a) \,:\, a \in \Sigma\} = \{r_1(u), \ldots, r_n(u)\}, \tag{4.200}$$

where the equality considers the two sides of the equation to be multisets, and moreover

$$r_1(u) \geq \cdots \geq r_n(u). \tag{4.201}$$

Theorem 4.30. *Let Σ be an alphabet and let $u, v \in \mathbb{R}^{\Sigma}$. The following statements are equivalent:*

1. $v \prec u$.
2. *For $n = |\Sigma|$, one has*

$$r_1(u) + \cdots + r_m(u) \geq r_1(v) + \cdots + r_m(v) \tag{4.202}$$

   for every choice of $m \in \{1, \ldots, n-1\}$, as well as

$$r_1(u) + \cdots + r_n(u) = r_1(v) + \cdots + r_n(v). \tag{4.203}$$

3. *There exists a unitary operator $U \in \mathrm{U}(\mathbb{C}^{\Sigma})$ such that, for the doubly stochastic operator $A \in \mathrm{L}(\mathbb{R}^{\Sigma})$ defined as*

$$A(a,b) = |U(a,b)|^2 \tag{4.204}$$

   for each $(a,b) \in \Sigma \times \Sigma$, one has $v = Au$.

Proof. Assume first that statement 1 holds, so that there exists a doubly stochastic operator $A \in \mathrm{L}(\mathbb{R}^{\Sigma})$ such that $Au = v$. It will be proved that

$$\sum_{a \in \Sigma} u(a) = \sum_{a \in \Sigma} v(a), \tag{4.205}$$

and that, for every subset $S \subseteq \Sigma$, there exists a subset $T \subseteq \Sigma$ such that $|S| = |T|$ and

$$\sum_{a \in T} u(a) \geq \sum_{a \in S} v(a). \tag{4.206}$$

This will imply statement 2; the condition (4.205) is equivalent to (4.203), while (4.206) implies (4.202) when one considers the case that S comprises the indices of the m largest entries of v, for each $m \in \{1, \ldots, n-1\}$. The first condition (4.205) is immediate from the assumption that A is stochastic:

$$\sum_{a \in \Sigma} v(a) = \sum_{a \in \Sigma} (Au)(a) = \sum_{a,b \in \Sigma} A(a,b)u(b) = \sum_{b \in \Sigma} u(b). \tag{4.207}$$

To prove the second condition, observe first that the Birkhoff–von Neumann theorem (Theorem 4.28) implies that

$$A = \sum_{\pi \in \mathrm{Sym}(\Sigma)} p(\pi) V_\pi \tag{4.208}$$

for some choice of a probability vector $p \in \mathcal{P}(\mathrm{Sym}(\Sigma))$. For an arbitrary choice of a subset $S \subseteq \Sigma$, the expression (4.208) implies that

$$\sum_{a \in S} v(a) = \sum_{a \in S} (Au)(a) = \sum_{\pi \in \mathrm{Sym}(\Sigma)} p(\pi) \sum_{a \in S} u\big(\pi^{-1}(a)\big) \tag{4.209}$$

A convex combination of a collection of real numbers cannot exceed the maximal element in that set, and therefore there must exist a permutation $\pi \in \mathrm{Sym}(\Sigma)$ such that

$$\sum_{a \in T_\pi} u(a) = \sum_{a \in S} u\big(\pi^{-1}(a)\big) \geq \sum_{a \in S} v(a) \tag{4.210}$$

for $T_\pi = \big\{\pi^{-1}(a) \,:\, a \in S\big\}$. As $|T_\pi| = |S|$, the inequality (4.206) has been proved for a suitable choice of an index set T. It has therefore been proved that statement 1 implies statement 2.

Next it will be proved that statement 2 implies statement 3, which is the most difficult implication of the proof. The implication will be proved by induction on $n = |\Sigma|$, for which the base case $n = 1$ is trivial. It will therefore be assumed that $n \geq 2$ for the remainder of the proof. As the majorization relationship is invariant under renaming and independently reordering the indices of the vectors under consideration, there is no loss of generality in assuming that $\Sigma = \{1, \ldots, n\}$, that $u = (u_1, \ldots, u_n)$ satisfies $u_1 \geq \cdots \geq u_n$, and that $v = (v_1, \ldots, v_n)$ satisfies $v_1 \geq \cdots \geq v_n$.

Under the assumption that statement 2 holds, it must be the case that $u_1 \geq v_1 \geq u_k$ for some choice of $k \in \{1, \ldots, n\}$. Fix k to be minimal among all such indices. There are two cases: $k = 1$ and $k > 1$.

If it is the case that $k = 1$, then $u_1 = v_1$, from which it follows that

$$u_2 + \cdots + u_m \geq v_2 + \cdots + v_m \tag{4.211}$$

for every $m \in \{2, \ldots, n-1\}$, as well as

$$u_2 + \cdots + u_n = v_2 + \cdots + v_n. \tag{4.212}$$

Define vectors $x = (u_2, \ldots, u_n)$ and $y = (v_2, \ldots, v_n)$. By the hypothesis of induction, there must therefore exist a unitary operator V, whose entries are indexed by the set $\{2, \ldots, n\}$, having the property that the doubly stochastic operator B defined by

$$B(a, b) = |V(a, b)|^2 \tag{4.213}$$

for all $a, b \in \{2, \ldots, n\}$ satisfies $y = Bx$. Taking U to be the unitary operator

$$U = \begin{pmatrix} 1 & 0 \\ 0 & V \end{pmatrix} \tag{4.214}$$

and letting A be defined by

$$A(a, b) = |U(a, b)|^2 \tag{4.215}$$

for all $a, b \in \{1, \ldots, n\}$, one has that $v = Au$, as required.

If it is the case that $k > 1$, then $u_1 > v_1 \geq u_k$, and so there must exist a real number $\lambda \in [0, 1)$ such that $v_1 = \lambda u_1 + (1 - \lambda) u_k$. Define vectors $x = (x_2, \ldots, x_n)$ and $y = (y_2, \ldots, y_n)$ as

$$\begin{aligned} &x = (u_2, \ldots, u_{k-1}, (1 - \lambda) u_1 + \lambda u_k, u_{k+1}, \ldots, u_n), \\ &y = (v_2, \ldots, v_n). \end{aligned} \tag{4.216}$$

For $m \in \{2, \ldots, k-1\}$ it holds that

$$x_2 + \cdots + x_m = u_2 + \cdots + u_m > (m-1) v_1 \geq v_2 + \cdots + v_m, \tag{4.217}$$

by virtue of the fact that k is the minimal index for which $v_1 \geq u_k$. For $m \in \{k, \ldots, n\}$ it holds that

$$\begin{aligned} &x_2 + \cdots + x_m \\ &\quad = (1 - \lambda) u_1 + u_2 + \cdots + u_{k-1} + \lambda u_k + u_{k+1} + \cdots + u_m \\ &\quad = u_1 + \cdots + u_m - v_1 \geq v_1 + \cdots + v_m - v_1 = v_2 + \cdots + v_m. \end{aligned} \tag{4.218}$$

By the hypothesis of induction, there must therefore exist a unitary operator V, whose entries are indexed by the set $\{2,\ldots,n\}$, having the property that the doubly stochastic operator B defined by

$$B(a,b) = |V(a,b)|^2 \tag{4.219}$$

for every $a,b \in \{2,\ldots,n\}$ satisfies $y = Bx$. Let W be the unitary operator defined by

$$\begin{aligned} We_1 &= \sqrt{\lambda}e_1 - \sqrt{1-\lambda}e_k, \\ We_k &= \sqrt{1-\lambda}e_1 + \sqrt{\lambda}e_k, \end{aligned} \tag{4.220}$$

and $We_a = e_a$ for $a \in \{2,\ldots,n\}\backslash\{k\}$, and let

$$U = \begin{pmatrix} 1 & 0 \\ 0 & V \end{pmatrix} W. \tag{4.221}$$

The entries of U may be calculated explicitly—they are given by

$$\begin{aligned} U(1,1) &= \sqrt{\lambda} & U(a,1) &= -\sqrt{1-\lambda}V(a,k) \\ U(1,k) &= \sqrt{1-\lambda} & U(a,k) &= \sqrt{\lambda}V(a,k) \\ U(1,b) &= 0 & U(a,b) &= V(a,b) \end{aligned} \tag{4.222}$$

for $a \in \{2,\ldots,n\}$ and $b \in \{2,\ldots,n\}\backslash\{k\}$. Letting A be the doubly stochastic operator defined by

$$A(a,b) = |U(a,b)|^2 \tag{4.223}$$

for every $a,b \in \{1,\ldots,n\}$, one obtains an operator whose entries are given by

$$\begin{aligned} A(1,1) &= \lambda & A(a,1) &= (1-\lambda)B(a,k) \\ A(1,k) &= 1-\lambda & A(a,k) &= \lambda B(a,k) \\ A(1,b) &= 0 & A(a,b) &= B(a,b) \end{aligned} \tag{4.224}$$

for $a \in \{2,\ldots,n\}$ and $b \in \{2,\ldots,n\}\backslash\{k\}$. Equivalently,

$$A = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix} D, \tag{4.225}$$

for D being the doubly stochastic operator defined by

$$\begin{aligned} De_1 &= \lambda e_1 + (1-\lambda)e_k, \\ De_k &= (1-\lambda)e_1 + \lambda e_k, \end{aligned} \tag{4.226}$$

and $De_a = e_a$ for $a \in \{2, \ldots, m\} \backslash \{k\}$. It holds that

$$Du = \begin{pmatrix} v_1 \\ x \end{pmatrix} \tag{4.227}$$

and therefore

$$Au = \begin{pmatrix} v_1 \\ Bx \end{pmatrix} = v. \tag{4.228}$$

It has therefore been proved that statement 2 implies statement 3.

The final step is to observe that statement 3 implies statement 1, which is trivial, as the operator A determined by statement 3 must be doubly stochastic. □

Remark 4.31. In light of the equivalence between the first and third statements in Theorem 4.30, it is natural to ask if every doubly stochastic operator $A \in \mathrm{L}(\mathbb{R}^\Sigma)$ is given by $A(a,b) = |U(a,b)|^2$ for some choice of a unitary operator $U \in \mathrm{U}(\mathbb{C}^\Sigma)$. This is not the case: the operator

$$A = \frac{1}{2} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \tag{4.229}$$

in $\mathrm{L}(\mathbb{R}^3)$ is an example of a doubly stochastic operator that cannot be derived from a unitary operator in this fashion. Indeed, if A is to be derived from a unitary operator $U \in \mathrm{U}(\mathbb{C}^3)$, then U must take the form

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & \alpha_2 & \alpha_1 \\ \alpha_3 & 0 & \beta_1 \\ \beta_3 & \beta_2 & 0 \end{pmatrix} \tag{4.230}$$

for $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$, and β_3 being complex numbers on the unit circle. However, if U is unitary, then it must hold that

$$\mathbb{1} = UU^* = \frac{1}{2} \begin{pmatrix} |\alpha_1|^2 + |\alpha_2|^2 & \alpha_1\overline{\beta_1} & \alpha_2\overline{\beta_2} \\ \overline{\alpha_1}\beta_1 & |\alpha_3|^2 + |\beta_1|^2 & \alpha_3\overline{\beta_3} \\ \overline{\alpha_2}\beta_2 & \overline{\alpha_3}\beta_3 & |\beta_2|^2 + |\beta_3|^2 \end{pmatrix}. \tag{4.231}$$

This, however, is impossible, as none of the off-diagonal entries of the operator on the right-hand-side of (4.231) can equal zero for $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$, and β_3 being complex numbers on the unit circle.

4.3.2 Majorization for Hermitian operators

The majorization relation for Hermitian operators will now be defined. This relation inherits the essential characteristics of its real vector analogue; and similar to its real vector analogue it may be characterized in multiple ways. After a discussion of its alternative characterizations, two applications of majorization for Hermitian operators will be presented.

Definition and characterizations of majorization for Hermitian operators

In analogy to the intuitive description of the majorization relation for real vectors suggested previously, one may view that one Hermitian operator X majorizes another Hermitian operator Y if and only if it is the case that Y can be obtained from X through a "random mixing" process. One natural way to formalize the notion of "random mixing" in the quantum setting is to consider mixed-unitary channels to be representative of such processes. The definition of the majorization relation for Hermitian operators that follows adopts this viewpoint.

Definition 4.32. Let $X, Y \in \mathrm{Herm}(\mathcal{X})$ be Hermitian operators, for $\mathcal{X}$ being a complex Euclidean space. It is said that X *majorizes* Y, written $Y \prec X$, if and only if there exists a mixed-unitary channel $\Phi \in \mathrm{C}(\mathcal{X})$ for which $\Phi(X) = Y$.

There is, *a priori*, no reason to prefer Definition 4.32 over one possible alternative, in which the condition that Φ is mixed-unitary is replaced by the condition that Φ is a unital channel. This is indeed a natural alternative because unital channels are, in some sense, analogous to doubly stochastic operators acting on real Euclidean spaces, while mixed-unitary channels are analogous to convex combinations of permutation operators. The failure of a direct quantum analogue to the Birkhoff–von Neumann theorem to hold is responsible for this apparent difference between two possible definitions of majorization for Hermitian operators.

The following theorem demonstrates that these two alternatives are, in fact, equivalent. The theorem also provides two additional characterizations of the majorization relation for Hermitian operators, the second of which establishes a direct link between majorization for Hermitian operators and majorization for real vectors (applied to the vectors of eigenvalues of the Hermitian operators under consideration).

Theorem 4.33 (Uhlmann). *Let $X, Y \in \mathrm{Herm}(\mathcal{X})$ be Hermitian operators, for $\mathcal{X}$ being a complex Euclidean space. The following statements are equivalent:*

1. $Y \prec X$.
2. *There exists a unital channel $\Phi \in \mathrm{C}(\mathcal{X})$ such that $Y = \Phi(X)$.*
3. *There exists a positive, trace-preserving, and unital map $\Phi \in \mathrm{T}(\mathcal{X})$ such that $Y = \Phi(X)$.*
4. $\lambda(Y) \prec \lambda(X)$.

Proof. Under the assumption that statement 1 holds, there exists a mixed-unitary channel $\Phi \in \mathrm{C}(\mathcal{X})$ such that $Y = \Phi(X)$. Any such channel is necessarily unital, and therefore statement 1 trivially implies statement 2. As every unital channel is positive, trace-preserving, and unital, statement 2 trivially implies statement 3.

Now assume that statement 3 holds. Let $n = \dim(\mathcal{X})$, and let

$$X = \sum_{j=1}^{n} \lambda_j(X)\, x_j x_j^* \qquad \text{and} \qquad Y = \sum_{k=1}^{n} \lambda_k(Y)\, y_k y_k^* \tag{4.232}$$

be spectral decompositions of X and Y, respectively. As $\Phi(X) = Y$, one concludes that

$$\lambda_k(Y) = \sum_{j=1}^{n} \lambda_j(X)\, y_k^* \Phi(x_j x_j^*) y_k \tag{4.233}$$

for each $k \in \{1, \ldots, n\}$. Equivalently, $\lambda(Y) = A\lambda(X)$ for $A \in \mathrm{L}(\mathbb{R}^n)$ being the operator defined as

$$A(k, j) = y_k^* \Phi(x_j x_j^*) y_k \tag{4.234}$$

for every $j, k \in \{1, \ldots, n\}$. Each entry of A is nonnegative by the positivity of Φ; by the fact that Φ is trace-preserving, it holds that

$$\sum_{k=1}^{n} A(k, j) = 1 \tag{4.235}$$

for each $j \in \{1, \ldots, n\}$; and by the fact that Φ is unital, it holds that

$$\sum_{j=1}^{n} A(k, j) = 1 \tag{4.236}$$

for each $k \in \{1, \ldots, n\}$. The operator A is therefore doubly stochastic, so that $\lambda(Y) \prec \lambda(X)$. It has therefore been proved that statement 3 implies statement 4.

Finally, assume $\lambda(Y) \prec \lambda(X)$, and again consider spectral decompositions of X and Y as in (4.232). As $\lambda(Y) \prec \lambda(X)$, one may conclude from Theorem 4.28 that there exists a probability vector $p \in \mathcal{P}(S_n)$ such that

$$\lambda_k(Y) = \sum_{\pi \in S_n} p(\pi) \lambda_{\pi(k)}(X) \tag{4.237}$$

for all $k \in \{1, \ldots, n\}$. By defining a unitary operator

$$U_\pi = \sum_{k=1}^{n} y_k x_{\pi(k)}^* \tag{4.238}$$

for each permutation $\pi \in S_n$, one has that

$$\begin{aligned} &\sum_{\pi \in S_n} p(\pi) U_\pi X U_\pi^* \\ &\quad = \sum_{k=1}^{n} \sum_{\pi \in S_n} p(\pi) \lambda_{\pi(k)}(X) y_k y_k^* = \sum_{k=1}^{n} \lambda_k(Y) y_k y_k^* = Y. \end{aligned} \tag{4.239}$$

It therefore holds that $Y \prec X$, and so statement 4 implies statement 1, which completes the proof. □

Two applications of Hermitian operator majorization

The theorems that follow offer a sample of the applications of majorization for Hermitian operators. The first theorem, whose proof makes essential use of Theorem 4.33, provides a precise characterization of those real vectors that may be obtained as the diagonal entries of a given Hermitian operator with respect to an arbitrary choice of an orthonormal basis.

Theorem 4.34 (Schur–Horn theorem)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let $n = \dim(\mathcal{X})$, and let $X \in \mathrm{Herm}(\mathcal{X})$ be a Hermitian operator. The following two implications, which are converse to one another, hold:*

1. *For every orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$, the vector $v \in \mathbb{R}^n$ defined by $v(k) = x_k^* X x_k$ for each $k \in \{1, \ldots, n\}$ satisfies $v \prec \lambda(X)$.*
2. *For every vector $v \in \mathbb{R}^n$ satisfying $v \prec \lambda(X)$, there exists an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$ for which $v(k) = x_k^* X x_k$ for each $k \in \{1, \ldots, n\}$.*

Proof. Suppose $\{x_1, \ldots, x_n\}$ is an orthonormal basis of $\mathcal{X}$ and $v \in \mathbb{R}^n$ is defined as $v(k) = x_k^* X x_k$ for each $k \in \{1, \ldots, n\}$. Define a map $\Phi \in \mathrm{T}(\mathcal{X})$ as

$$\Phi(Y) = \sum_{k=1}^{n} x_k x_k^* Y x_k x_k^* \tag{4.240}$$

for every operator $Y \in \mathrm{L}(\mathcal{X})$, and observe that Φ is a pinching channel. By Proposition 4.6, it follows that Φ is a mixed-unitary channel. One therefore has $\Phi(X) \prec X$, which implies $\lambda(\Phi(X)) \prec \lambda(X)$ by Theorem 4.33. As

$$\Phi(X) = \sum_{k=1}^{n} v(k) x_k x_k^*, \tag{4.241}$$

it is evident that

$$\operatorname{spec}(\Phi(X)) = \{v(1), \ldots, v(n)\}, \tag{4.242}$$

or equivalently that

$$\lambda(\Phi(X)) = W_\pi v \tag{4.243}$$

for a permutation operator W_π that has the effect of ordering the entries of v from largest to smallest:

$$(W_\pi v)(1) \geq \cdots \geq (W_\pi v)(n). \tag{4.244}$$

It follows that $v \prec \lambda(X)$, as is required to establish the first implication.

Now suppose $v \in \mathbb{R}^n$ is a vector satisfying $v \prec \lambda(X)$, and let

$$X = \sum_{k=1}^{n} \lambda_k(X) u_k u_k^* \tag{4.245}$$

be a spectral decomposition of X. By Theorem 4.30, the assumption that $v \prec \lambda(X)$ implies that there exists a unitary operator $U \in \mathrm{U}(\mathbb{C}^n)$ such that, for $A \in \mathrm{L}(\mathbb{R}^n)$ defined by

$$A(j,k) = |U(j,k)|^2 \tag{4.246}$$

for $j, k \in \{1, \ldots, n\}$, one has $v = A\lambda(X)$. Define $V \in \mathrm{U}(\mathcal{X}, \mathbb{C}^n)$ as

$$V = \sum_{k=1}^{n} e_k u_k^* \tag{4.247}$$

and let

$$x_k = V^* U^* V u_k \tag{4.248}$$

for each $k \in \{1, \ldots, n\}$. The operator $V^* U^* V \in \mathrm{U}(\mathcal{X})$ is a unitary operator, implying that $\{x_1, \ldots, x_n\}$ is an orthonormal basis of $\mathcal{X}$. It holds that

$$x_k^* X x_k = \sum_{j=1}^{n} |U(k,j)|^2 \lambda_j(X) = (A\lambda(X))(k) = v(k), \tag{4.249}$$

which establishes the second implication. □

The next theorem, representing a second application of majorization for Hermitian operators, characterizes the collection of probability vectors that are consistent with the representation of a given density operator as a mixture of pure states.

Theorem 4.35. *Let $\mathcal{X}$ be a complex Euclidean space, let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, let $n = \dim(\mathcal{X})$, and let $p = (p_1, \ldots, p_n)$ be a probability vector. There exist a collection of (not necessarily orthogonal) unit vectors $\{u_1, \ldots, u_n\} \subset \mathcal{X}$ such that*

$$\rho = \sum_{k=1}^{n} p_k u_k u_k^* \tag{4.250}$$

if and only if $p \prec \lambda(\rho)$.

Proof. Assume first that

$$\rho = \sum_{k=1}^{n} p_k u_k u_k^* \tag{4.251}$$

for a collection $\{u_1, \ldots, u_n\} \subset \mathcal{X}$ of unit vectors. Define $A \in \mathrm{L}(\mathbb{C}^n, \mathcal{X})$ as

$$A = \sum_{k=1}^{n} \sqrt{p_k}\, u_k e_k^*, \tag{4.252}$$

and observe that $AA^* = \rho$. It holds that

$$A^* A = \sum_{j=1}^{n} \sum_{k=1}^{n} \sqrt{p_j p_k}\, \langle u_k, u_j \rangle\, E_{k,j}\,, \tag{4.253}$$

and therefore

$$e_k^* A^* A e_k = p_k \tag{4.254}$$

for every $k \in \{1, \ldots, n\}$. By Theorem 4.34, this implies $p \prec \lambda(A^* A)$. As

$$\lambda(A^* A) = \lambda(AA^*) = \lambda(\rho), \tag{4.255}$$

it follows that $p \prec \lambda(\rho)$. One of the required implications of the theorem has therefore been proved.

Now assume that $p \prec \lambda(\rho)$. By Theorem 4.34, there exists an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$ with the property that

$$p_k = x_k^* \rho x_k \tag{4.256}$$

for each $k \in \{1, \ldots, n\}$. Let

$$y_k = \sqrt{\rho} x_k \tag{4.257}$$

and define

$$u_k = \begin{cases} \frac{y_k}{\|y_k\|} & \text{if } y_k \neq 0 \\ z & \text{if } y_k = 0 \end{cases} \tag{4.258}$$

for each $k \in \{1, \ldots, n\}$, where $z \in \mathcal{X}$ is an arbitrarily chosen unit vector. One has that

$$\|y_k\|^2 = \left\langle \sqrt{\rho} x_k, \sqrt{\rho} x_k \right\rangle = x_k^* \rho x_k = p_k, \tag{4.259}$$

for each $k \in \{1, \ldots, n\}$, and therefore

$$\sum_{k=1}^{n} p_k u_k u_k^* = \sum_{k=1}^{n} y_k y_k^* = \sum_{k=1}^{n} \sqrt{\rho} x_k x_k^* \sqrt{\rho} = \rho. \tag{4.260}$$

This proves the other required implication of the theorem. □

4.4 Exercises

4.1. Let $\mathcal{X}$ be a complex Euclidean space with $\dim(\mathcal{X}) = 3$ and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a Schur channel. Prove that Φ is a mixed-unitary channel.

4.2. For every positive integer $n \geq 2$, define a unital channel $\Phi_n \in \mathrm{C}(\mathbb{C}^n)$ as

$$\Phi_n(X) = \frac{\mathrm{Tr}(X)\mathbb{1}_n - X^{\mathsf{T}}}{n-1} \tag{4.261}$$

for every $X \in \mathrm{L}(\mathbb{C}^n)$, where $\mathbb{1}_n$ denotes the identity operator on $\mathbb{C}^n$. Prove that Φ_n is not mixed-unitary when n is odd.

A correct solution to this exercise generalizes Example 4.3, but a different argument will be needed than the one in that example when $n \geq 5$.

4.3. Let n be a positive integer, let $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$, let $\{W_{a,b} : a, b \in \mathbb{Z}_n\} \subset \mathrm{U}(\mathcal{X})$ be the set of discrete Weyl operators acting on $\mathcal{X}$, and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a channel. Prove that the following two statements are equivalent:

1. Φ is both a Schur channel and a Weyl-covariant channel.
2. There exists a probability vector $p \in \mathcal{P}(\mathbb{Z}_n)$ such that

$$\Phi(X) = \sum_{a \in \mathbb{Z}_n} p(a) W_{0,a} X W_{0,a}^* \tag{4.262}$$

   for all $X \in \mathrm{L}(\mathcal{X})$.

4.4. Let $\mathcal{X}$ be a complex Euclidean space and let $\Phi \in \mathrm{T}(\mathcal{X})$ be a Hermiticity-preserving map. Prove that the following two statements are equivalent:

1. Φ is positive, trace-preserving, and unital.
2. $\Phi(H) \prec H$ for every Hermitian operator $H \in \mathrm{Herm}(\mathcal{X})$.

4.5. Let $\mathcal{X}$ be a complex Euclidean space, let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, let $p = (p_1, \ldots, p_m)$ be a probability vector, and assume $p_1 \geq p_2 \geq \cdots \geq p_m$. Prove that there exist unit vectors $u_1, \ldots, u_m \in \mathcal{X}$ satisfying

$$\rho = \sum_{k=1}^{m} p_k u_k u_k^* \tag{4.263}$$

if and only if

$$p_1 + \cdots + p_k \leq \lambda_1(\rho) + \cdots + \lambda_k(\rho) \tag{4.264}$$

for all k satisfying $1 \leq k \leq \mathrm{rank}(\rho)$.

A correct solution to this problem generalizes Theorem 4.35, as m need not coincide with the dimension of $\mathcal{X}$.

4.6. Let $\mathcal{X}$ be a complex Euclidean space, let $n = \dim(\mathcal{X})$, and let $\Phi \in \mathrm{C}(\mathcal{X})$ be a unital channel. Following the conventions discussed in Section 1.1.3 of Chapter 1, let $s_1(Y) \geq \cdots \geq s_n(Y)$ denote the singular values of a given operator $Y \in \mathrm{L}(\mathcal{X})$, ordered from largest to smallest, and taking $s_k(Y) = 0$ when $k > \mathrm{rank}(Y)$. Prove that, for every operator $X \in \mathrm{L}(\mathcal{X})$, it holds that

$$s_1(X) + \cdots + s_m(X) \geq s_1(\Phi(X)) + \cdots + s_m(\Phi(X)) \tag{4.265}$$

for every $m \in \{1, \ldots, n\}$.

4.5 Bibliographic remarks

Unital channels are sometimes referred to as *doubly stochastic* maps in the mathematics literature, although that term has also been used in reference to positive (but not necessarily completely positive), trace-preserving, and unital maps. The extreme points of sets of unital channels were studied by Landau and Streater [142]; the facts represented by Theorem 4.21, Example 4.3, and Theorem 4.23 appear in their paper. Related results for positive, trace-preserving, and unital maps had previously been discovered by Tregub [202], who also gave a different example of a unital channel (that also happens to be a Schur channel) that is not mixed-unitary. Another class of examples of this type appear in the work of Kümmerer and Maassen [140].

Mixed-unitary channels have often been called *random unitary* channels, as in the case of Audenaert and Scheel [18]. One disadvantage of this terminology is that it clashes with the terminology associated with a different concept, which is the consideration of unitary operators chosen at random from a given distribution (often the distribution associated with the Haar measure, as will be discussed later in Chapter 7).

The notion of environment-assisted channel correction was suggested by Alber, Beth, Charnes, Delgado, Grassl, and Mussinger [4]. The characterization of mixed-unitary channels based on this notion, as established by Theorem 4.8, follows from a slightly more general result due to Gregoratti and Werner [80]. Corollary 4.11 was proved by Buscemi [43] through the use of this characterization together with Corollary 2.48.

The discrete Weyl operators appear in Weyl's work on group-theoretic aspects of quantum mechanics. (See, for instance, Sections 14 and 15 in Chapter IV of [228].) The notion of covariance applies not only to the discrete Weyl operators and quantum channels, but to other collections of unitary operators and algebraic objects. There is, for example, some discussion of this notion in [228], and it was considered more explicitly for quantum instruments by Davies [54]. Channel covariance with respect to the discrete Weyl operators was considered by Holevo [109, 110], and the facts represented by Theorem 4.14 may be derived from that work.

Schur [190] proved that the positive semidefinite cone is closed under entry-wise products—which is a fact now referred to as the *Schur product theorem*. The entry-wise product of operators is called the *Schur product*, and Schur maps are so named for this reason. The term *Hadamard product* is

also sometimes used to refer to the entry-wise product, and correspondingly Schur maps are sometimes referred to as *Hadamard maps*. Schur maps are also referred to as *diagonal maps* by some authors, as they correspond to maps with diagonal Kraus operators (as is stated in Theorem 4.19).

Theorem 4.25 is due to Kribs [138], whose proof made use of arguments that can be found in the paper of Lindblad [149]. Fixed points of quantum channels, unital channels, and other classes of completely positive maps have also been studied by other researchers, including Bratteli, Jorgensen, Kishimoto, and Werner [42], Arias, Gheondea, and Gutter [13], and others. Theorem 4.27 is a special case of a theorem due to Perez-García, Wolf, Petz, and Ruskai [172]. (The theorem holds for a more general class of norms, not just the spectral norm.)

The notion of majorization for real vectors was developed in the first half of the twentieth century by several mathematicians, including Hardy, Littlewood, Pólya, Schur, Rado, and Horn. Details on this history may be found in Marshall, Olkin, and Arnold [154]. The extension of this notion to Hermitian operators is due to Uhlmann [206, 207, 208], as is Theorem 4.33. (See also the book of Alberti and Uhlmann [6].) The two implications of Theorem 4.34 were proved by Schur [189] and Horn [113], respectively, and Theorem 4.35 is due to Nielsen [164].

Chapter 5

Quantum entropy and source coding

The *von Neumann entropy* of a quantum state is an information-theoretic measure of the amount of randomness or uncertainty that is inherent to that state, and the *quantum relative entropy* of one quantum state with respect to another is a related measure of the degree to which the first state differs from the second. This chapter defines these function, establishes some of their fundamental properties, and explains their connections to the task of *source coding*.

5.1 Classical entropy

The von Neumann entropy and quantum relative entropy functions are quantum analogues of classical information-theoretic notions: the Shannon entropy and (classical) relative entropy functions. It is appropriate to begin the chapter with a discussion of these classical notions, as an investigation of the mathematical properties of the von Neumann entropy and quantum relative entropy functions builds naturally on their classical counterparts.

5.1.1 Definitions of classical entropic functions

With respect to the definition that follows, the Shannon entropy is specified for every vector with nonnegative entries, over any real Euclidean space. Although it is most common that this function is considered in the case that

its argument is a probability vector, it is convenient nevertheless to extend its domain in this way.

Definition 5.1. Let Σ be an alphabet and let $u \in [0,\infty)^{\Sigma}$ be a vector of non-negative real numbers indexed by Σ. One defines the *Shannon entropy* of the vector u as

$$\mathrm{H}(u) = -\sum_{\substack{a\in\Sigma \\ u(a)>0}} u(a)\log(u(a)). \tag{5.1}$$

The Shannon entropy $\mathrm{H}(p)$ of a probability vector $p \in \mathcal{P}(\Sigma)$ is sometimes described as the amount of randomness inherent to the distribution described by p, measured in bits. Alternatively, $\mathrm{H}(p)$ may be described as the number of bits of uncertainty one has regarding the outcome of a random process described by p before the outcome is learned, or as the number of bits of information one gains as a result of learning which element $a \in \Sigma$ has been produced by such a process.

In the simple case that $\Sigma = \{0,1\}$ and $p(0) = p(1) = 1/2$, for instance, it holds that $\mathrm{H}(p) = 1$. This is natural, as one would expect that the amount of uncertainty of a uniformly generated random bit, measured in bits, would be 1 bit of uncertainty. In contrast, for a completely deterministic process, meaning one in which p is an elementary unit vector, there is no randomness or uncertainty, and no information gain when the selection is learned. Correspondingly, one has that the entropy $\mathrm{H}(p)$ is zero in this case.

It is important to recognize, however, that intuitive descriptions of the Shannon entropy, as a measure of randomness, uncertainty, or information gain, must be viewed as representing *expectations* rather than absolute or definitive measures. The following example illustrates this point.

Example 5.2. Let m be a positive integer, let

$$\Sigma = \left\{0, 1, \ldots, 2^{m^2}\right\}, \tag{5.2}$$

and define a probability vector $p \in \mathcal{P}(\Sigma)$ as follows:

$$p(a) = \begin{cases} 1 - \frac{1}{m} & \text{if } a = 0 \\ \frac{1}{m} 2^{-m^2} & \text{if } 1 \leq a \leq 2^{m^2}. \end{cases} \tag{5.3}$$

A calculation reveals that $\mathrm{H}(p) > m$, and yet the outcome 0 appears with probability $1 - 1/m$ in a random selection described by p. So, as m grows,

one becomes more and more "certain" that the outcome will be 0, and yet the "uncertainty" (as measured by the entropy) increases.

This example does not represent a paradox or suggest that the Shannon entropy is not reasonably viewed as a measure of uncertainty. If one considers an experiment in which a very large number of elements of Σ are selected independently, each according to the probability vector p, then the value $\mathrm{H}(p)$ indeed does correspond more intuitively to the average or expected amount of uncertainty of each random selection.

Sometimes one speaks of the Shannon entropy of a classical register X, with the notation $\mathrm{H}(\mathsf{X})$ being used for this purpose. This is a convenient shorthand to be interpreted as meaning $\mathrm{H}(p)$, for the probability vector p describing the probabilistic state of X at the moment under consideration. Notations such as $\mathrm{H}(\mathsf{X},\mathsf{Y})$ and $\mathrm{H}(\mathsf{X}_1,\ldots,\mathsf{X}_n)$ are used in place of $\mathrm{H}((\mathsf{X},\mathsf{Y}))$ and $\mathrm{H}((\mathsf{X}_1,\ldots,\mathsf{X}_n))$ when referring to the Shannon entropy of compound registers. Along similar lines, the notation $\mathrm{H}(\alpha_1,\ldots,\alpha_n)$ will be used in place of $\mathrm{H}((\alpha_1,\ldots,\alpha_n))$ when it is convenient to refer to the entropy of a vector written as $(\alpha_1,\ldots,\alpha_n)$.

The *relative entropy* function, which is also known as the *Kullback–Leibler divergence*, is closely related to the Shannon entropy. For the purposes of this book, the primary motivation for its introduction is that it serves as a useful analytic tool for reasoning about the Shannon entropy.

Definition 5.3. Let Σ be an alphabet and let $u, v \in [0,\infty)^{\Sigma}$ be vectors of nonnegative real numbers indexed by Σ. The *relative entropy* $\mathrm{D}(u\|v)$ of u with respect to v is defined as follows. If it is the case that $\operatorname{supp}(u) \subseteq \operatorname{supp}(v)$ (i.e., $u(a) > 0$ implies $v(a) > 0$ for all $a \in \Sigma$), then $\mathrm{D}(u\|v)$ is defined as

$$\mathrm{D}(u\|v) = \sum_{\substack{a\in\Sigma \\ u(a)>0}} u(a)\log\left(\frac{u(a)}{v(a)}\right). \tag{5.4}$$

For all other choices of u and v, one defines $\mathrm{D}(u\|v) = \infty$.

Like the Shannon entropy function, the relative entropy is most typically considered in cases where its arguments are probability vectors, but again it is convenient to extend its domain to arbitrary nonnegative real vectors.

For a given pair of probability vectors $p, q \in \mathcal{P}(\Sigma)$, the relative entropy $\mathrm{D}(p\|q)$ may be viewed as a measure of how much p differs from q in a

certain information-theoretic sense. Analytically speaking, it fails to satisfy the requirements of being a true metric: it is not symmetric, it takes infinite values for some pairs of inputs, and it does not satisfy the triangle inequality. When extended to arbitrary vectors of the form $u, v \in [0, \infty)^{\Sigma}$, it may also take negative values. Despite these apparent shortcomings, the relative entropy is an indispensable information-theoretic tool.

Two additional functions derived from the Shannon entropy function are the *conditional Shannon entropy* and the *mutual information*. Both concern correlations between two classical registers X and Y, and are functions of the joint probabilistic state of the pair (X, Y). The conditional Shannon entropy of X given Y is defined as

$$\mathrm{H}(\mathsf{X} \,|\, \mathsf{Y}) = \mathrm{H}(\mathsf{X}, \mathsf{Y}) - \mathrm{H}(\mathsf{Y}). \tag{5.5}$$

Intuitively speaking, this quantity represents the expected amount of uncertainty regarding the classical state of X one would have upon learning the classical state of Y. The *mutual information* between X and Y is defined as

$$\mathrm{I}(\mathsf{X} : \mathsf{Y}) = \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X}, \mathsf{Y}). \tag{5.6}$$

This quantity can alternately be expressed as

$$\mathrm{I}(\mathsf{X} : \mathsf{Y}) = \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{Y}|\mathsf{X}) = \mathrm{H}(\mathsf{X}) - \mathrm{H}(\mathsf{X}|\mathsf{Y}). \tag{5.7}$$

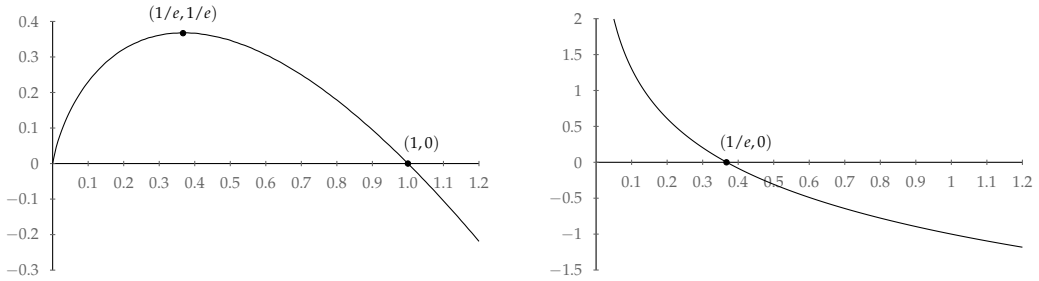
One typically views this quantity as representing the expected amount of information about X that one gains by learning the classical state of Y, or (equivalently) that one gains about Y by learning the classical state of X.

5.1.2 Properties of classical entropic functions

The Shannon and relative entropy functions possess a variety of useful and interesting properties. This section establishes several basic properties of these functions.

Scalar analogues of Shannon entropy and relative entropy

For the purposes of establishing basic analytic properties of the Shannon and relative entropy functions, it is helpful to define functions representing scalar analogues of these functions. These scalar functions are to be defined with respect to the natural logarithm rather than the base-2 logarithm, as



(a) $\eta(\alpha) = -\alpha \ln(\alpha)$ (b) $\eta'(\alpha) = -(1 + \ln(\alpha))$

Figure 5.1: Plots of the functions η and η'.

this will simplify some of the calculations to follow, particularly when they make use of differential calculus.

The first function $\eta : [0, \infty) \to \mathbb{R}$, which represents a scalar analogue of the Shannon entropy, is defined as follows:

$$\eta(\alpha) = \begin{cases} -\alpha \ln(\alpha) & \alpha > 0 \\ 0 & \alpha = 0. \end{cases} \tag{5.8}$$

The function η is continuous everywhere on its domain, and derivatives of η of all orders exist for all positive real numbers. In particular,

$$\eta'(\alpha) = -(1 + \ln(\alpha)) \tag{5.9}$$

and

$$\eta^{(n+1)}(\alpha) = \frac{(-1)^n}{\alpha^n} \tag{5.10}$$

for $n \geq 1$, for all $\alpha > 0$. Plots of the function η its first derivative η' are shown in Figure 5.1. As the second derivative of η is negative for all $\alpha > 0$, one has that η is a concave function:

$$\eta(\lambda\alpha + (1 - \lambda)\beta) \geq \lambda\eta(\alpha) + (1 - \lambda)\eta(\beta) \tag{5.11}$$

for all $\alpha, \beta \geq 0$ and $\lambda \in [0, 1]$.

The second function $\theta : [0, \infty)^2 \to (-\infty, \infty]$, which represents a scalar analogue of the relative entropy, is defined as follows:

$$\theta(\alpha, \beta) = \begin{cases} 0 & \text{if } \alpha = 0 \\ \infty & \text{if } \alpha > 0 \text{ and } \beta = 0 \\ \alpha \ln(\alpha) - \alpha \ln(\beta) & \text{if } \alpha > 0 \text{ and } \beta > 0. \end{cases} \tag{5.12}$$

It is evident from this definition that, when restricted to positive real number arguments $\alpha, \beta > 0$, the value $\theta(\alpha, \beta)$ is negative when $\alpha < \beta$, zero when $\alpha = \beta$, and positive when $\alpha > \beta$.

It is useful to note that the functions θ and η are related by the identity

$$\theta(\alpha, \beta) = -\beta \, \eta\left(\frac{\alpha}{\beta}\right), \tag{5.13}$$

which holds for all $\alpha \in [0, \infty)$ and $\beta \in (0, \infty)$. The function θ is continuous at every point (α, β) for which $\beta > 0$. It is not continuous at any point $(\alpha, 0)$, however, as every neighborhood of such a point contains both finite and infinite values.

The following useful lemma regarding the function θ is equivalent to a fact commonly known as the *log-sum inequality*.

Lemma 5.4. *Let $\alpha_0, \alpha_1, \beta_0, \beta_1 \in [0, \infty)$ be nonnegative real numbers. It holds that*

$$\theta(\alpha_0 + \alpha_1, \beta_0 + \beta_1) \leq \theta(\alpha_0, \beta_0) + \theta(\alpha_1, \beta_1). \tag{5.14}$$

Proof. If either of β_0 or β_1 is zero, the inequality is straightforward. More specifically, if $\beta_0 = 0$ and $\alpha_0 = 0$, the inequality is equivalent to

$$\theta(\alpha_1, \beta_1) \leq \theta(\alpha_1, \beta_1), \tag{5.15}$$

which is trivial, while if $\beta_0 = 0$ and $\alpha_0 > 0$, the right-hand side of (5.14) is infinite. A similar argument holds when $\beta_1 = 0$ by symmetry.

In the case that both β_0 and β_1 are positive, the inequality may be proved by combining the identity (5.13) with the concavity of η:

$$\begin{aligned}
&\theta(\alpha_0, \beta_0) + \theta(\alpha_1, \beta_1) \\
&\quad = -(\beta_0 + \beta_1) \left[\frac{\beta_0}{\beta_0 + \beta_1} \, \eta\left(\frac{\alpha_0}{\beta_0}\right) + \frac{\beta_1}{\beta_0 + \beta_1} \, \eta\left(\frac{\alpha_1}{\beta_1}\right) \right] \\
&\quad \geq -(\beta_0 + \beta_1) \, \eta\left(\frac{\alpha_0 + \alpha_1}{\beta_0 + \beta_1}\right) \\
&\quad = \theta(\alpha_0 + \alpha_1, \beta_0 + \beta_1),
\end{aligned} \tag{5.16}$$

as claimed. □

Elementary properties of Shannon entropy and relative entropy

The Shannon entropy function may be expressed in terms of the η-function as follows:

$$\mathrm{H}(u) = \frac{1}{\ln(2)} \sum_{a \in \Sigma} \eta(u(a)), \tag{5.17}$$

for every choice of an alphabet Σ and a vector $u \in [0, \infty)^{\Sigma}$. As the function η is continuous everywhere on its domain, the Shannon entropy function is continuous everywhere on its domain as well. The concavity of η implies the concavity of the Shannon entropy, as the following proposition states.

Proposition 5.5 (Concavity of Shannon entropy)**.** *Let Σ be an alphabet, let $u, v \in [0, \infty)^{\Sigma}$ be vectors, and let $\lambda \in [0, 1]$. It holds that*

$$\mathrm{H}(\lambda u + (1 - \lambda)v) \geq \lambda \, \mathrm{H}(u) + (1 - \lambda) \, \mathrm{H}(v). \tag{5.18}$$

Proof. By the concavity of the function η, one has

$$\begin{aligned} \mathrm{H}(\lambda u + (1 - \lambda)v) &= \frac{1}{\ln(2)} \sum_{a \in \Sigma} \eta(\lambda u(a) + (1 - \lambda)v(a)) \\ &\geq \frac{\lambda}{\ln(2)} \sum_{a \in \Sigma} \eta(u(a)) + \frac{1 - \lambda}{\ln(2)} \sum_{a \in \Sigma} \eta(v(a)) \\ &= \lambda \, \mathrm{H}(u) + (1 - \lambda) \, \mathrm{H}(v), \end{aligned} \tag{5.19}$$

as required. □

The next proposition states two identities that involve the Shannon entropy of direct sums and tensor products of vectors. Both identities may be verified through direct calculations.

Proposition 5.6. *Let Σ and Γ be alphabets and let $u \in [0, \infty)^{\Sigma}$ and $v \in [0, \infty)^{\Gamma}$ be vectors. It holds that*

$$\mathrm{H}(u \oplus v) = \mathrm{H}(u) + \mathrm{H}(v) \tag{5.20}$$

and

$$\mathrm{H}(u \otimes v) = \mathrm{H}(u) \sum_{b \in \Gamma} v(b) + \mathrm{H}(v) \sum_{a \in \Sigma} u(a). \tag{5.21}$$

One may observe that, for any choice of probability vectors $p \in \mathcal{P}(\Sigma)$ and $q \in \mathcal{P}(\Gamma)$, the identity (5.21) implies that

$$\mathrm{H}(p \otimes q) = \mathrm{H}(p) + \mathrm{H}(q). \tag{5.22}$$

As a special case of the same identity, one finds that

$$\mathrm{H}(\alpha p) = \alpha \, \mathrm{H}(p) - \alpha \log(\alpha) \tag{5.23}$$

for every scalar $\alpha > 0$ and every probability vector $p \in \mathcal{P}(\Sigma)$.

The relative entropy function may be expressed using the θ-function as follows:

$$\mathrm{D}(u\|v) = \frac{1}{\ln(2)} \sum_{a \in \Sigma} \theta(u(a), v(a)), \tag{5.24}$$

for every choice of an alphabet Σ and two vectors $u, v \in [0,\infty)^{\Sigma}$. It therefore holds that the relative entropy function is continuous when its domain is restricted to choices of v for which $\mathrm{supp}(v) = \Sigma$, but is not continuous at any point (u, v) for which $\mathrm{supp}(v) \neq \Sigma$.

The next proposition, which implies that the relative entropy between any two probability vectors is nonnegative, represents one application of Lemma 5.4.

Proposition 5.7. *Let Σ be an alphabet and let $u, v \in [0,\infty)^{\Sigma}$ be vectors. If it holds that*

$$\sum_{a \in \Sigma} u(a) \geq \sum_{a \in \Sigma} v(a), \tag{5.25}$$

then $\mathrm{D}(u\|v) \geq 0$. In particular, $\mathrm{D}(p\|q) \geq 0$ for all choices of probability vectors $p, q \in \mathcal{P}(\Sigma)$.

Proof. By Lemma 5.4, it holds that

$$\mathrm{D}(u\|v) = \frac{1}{\ln(2)} \sum_{a \in \Sigma} \theta(u(a), v(a)) \geq \frac{1}{\ln(2)} \, \theta\left(\sum_{a \in \Sigma} u(a), \sum_{a \in \Sigma} v(a) \right). \tag{5.26}$$

The proposition follows from the fact that $\theta(\alpha, \beta) \geq 0$ for every choice of nonnegative real numbers $\alpha, \beta \in [0,\infty)$ satisfying $\alpha \geq \beta$. □

Remark 5.8. Theorem 5.17 establishes a quantitative lower-bound on the relative entropy $\mathrm{D}(p\|q)$ in terms of the 1-norm distance $\|p - q\|_1$ between any two probability vectors p and q.

Proposition 5.7 may be used to prove upper and lower bounds on the Shannon entropy, as in the proof of the following proposition.

Proposition 5.9. *Let Σ be an alphabet, let $u \in [0,\infty)^{\Sigma}$ be a vector, and let*

$$\alpha = \sum_{a \in \Sigma} u(a). \tag{5.27}$$

It holds that

$$0 \leq \mathrm{H}(u) + \alpha \log(\alpha) \leq \alpha \log(|\Sigma|). \tag{5.28}$$

In particular, it holds that $0 \leq \mathrm{H}(p) \leq \log(|\Sigma|)$ for every probability vector $p \in \mathcal{P}(\Sigma)$.

Proof. First, suppose $p \in \mathcal{P}(\Sigma)$ is a probability vector. The Shannon entropy $\mathrm{H}(p)$ may be written as

$$\mathrm{H}(p) = \sum_{\substack{a \in \Sigma \\ p(a) > 0}} p(a) \log\left(\frac{1}{p(a)}\right), \tag{5.29}$$

which is a convex combination of nonnegative real numbers, by virtue of the fact that $p(a) \leq 1$ for each $a \in \Sigma$. It follows that $\mathrm{H}(p) \geq 0$.

Next, let $q \in \mathcal{P}(\Sigma)$ be the probability vector defined by $q(a) = 1/|\Sigma|$ for each $a \in \Sigma$. One may evaluate the relative entropy $\mathrm{D}(p\|q)$ directly from its definition, obtaining

$$\begin{aligned} \mathrm{D}(p\|q) &= \sum_{a \in \Sigma} p(a) \log(p(a)) - \sum_{a \in \Sigma} p(a) \log\left(\frac{1}{|\Sigma|}\right) \\ &= -\mathrm{H}(p) + \log(|\Sigma|). \end{aligned} \tag{5.30}$$

As p and q are probability vectors, Proposition 5.7 implies that the relative entropy $\mathrm{D}(p\|q)$ is nonnegative, and therefore $\mathrm{H}(p) \leq \log(|\Sigma|)$.

Now consider $u \in [0,\infty)^{\Sigma}$ and α, as in the statement of the proposition. If it is the case that $\alpha = 0$, then it must hold that u is the zero vector, in which case the proposition may be verified directly. Otherwise, let $p \in \mathcal{P}(\Sigma)$ be the probability vector defined by the equation $\alpha p = u$. By (5.23), one has

$$\mathrm{H}(u) = \mathrm{H}(\alpha p) = \alpha \, \mathrm{H}(p) - \alpha \log(\alpha). \tag{5.31}$$

Given that $0 \leq \mathrm{H}(p) \leq \log(|\Sigma|)$, it follows that

$$-\alpha \log(\alpha) \leq \mathrm{H}(u) \leq \alpha \log(|\Sigma|) - \alpha \log(\alpha), \tag{5.32}$$

which completes the proof. □

Proposition 5.7 also leads to a proof that the Shannon entropy is subadditive, in the sense described by the proposition that follows. Intuitively speaking, this property reflects the idea that the amount of uncertainty one has about a compound register cannot be greater than the total uncertainty one has about its individual registers.

Proposition 5.10 (Subadditivity of Shannon entropy)**.** *Let* X *and* Y *be classical registers. With respect to an arbitrary probabilistic state of these registers, it holds that*

$$\mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}). \tag{5.33}$$

Proof. Let $p \in \mathcal{P}(\Sigma \times \Gamma)$ denote an arbitrary probabilistic state of the pair (X,Y), for Σ and Γ being the classical state sets of X and Y, respectively. A calculation based on the definition of the relative entropy and elementary properties of logarithms reveals the equality

$$\mathrm{D}\bigl(p \bigl\| p[\mathsf{X}] \otimes p[\mathsf{Y}]\bigr) = \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}). \tag{5.34}$$

As the relative entropy of one probability vector with respect to another is nonnegative by Proposition 5.7, the required inequality follows. □

One may observe that Proposition 5.10 is equivalent to the statement that the mutual information $\mathrm{I}(\mathsf{X}:\mathsf{Y})$ between two registers is necessarily nonnegative, or equivalently that the conditional Shannon entropy $\mathrm{H}(\mathsf{Y}|\mathsf{X})$ of one register Y given another register X is no larger than the (unconditional) Shannon entropy $\mathrm{H}(\mathsf{Y})$ of the register Y alone: $\mathrm{H}(\mathsf{Y}|\mathsf{X}) \leq \mathrm{H}(\mathsf{Y})$.

The next proposition establishes a related fact: the Shannon entropy of a pair of classical registers (X,Y) cannot be less than the Shannon entropy of either of the registers viewed in isolation. Equivalently, the conditional Shannon entropy $\mathrm{H}(\mathsf{X}|\mathsf{Y})$ is nonnegative for all possible probabilistic states of the pair (X,Y).

Proposition 5.11. *Let* X *and* Y *be classical registers. With respect to an arbitrary probabilistic state of these registers, it holds that*

$$\mathrm{H}(\mathsf{X}) \leq \mathrm{H}(\mathsf{X},\mathsf{Y}). \tag{5.35}$$

Proof. Let Σ and Γ denote the classical state sets of X and Y, respectively, and let $p \in \mathcal{P}(\Sigma \times \Gamma)$ be an arbitrary probabilistic state of (X,Y). The logarithm

is an increasing function, and therefore

$$\log(p(a,b)) \leq \log\left(\sum_{c\in\Gamma} p(a,c)\right) \tag{5.36}$$

for every pair $(a,b) \in \Sigma \times \Gamma$. It follows that

$$\begin{aligned} \mathrm{H}(\mathsf{X},\mathsf{Y}) &= -\sum_{a\in\Sigma}\sum_{b\in\Gamma} p(a,b)\log(p(a,b)) \\ &\geq -\sum_{a\in\Sigma}\left(\sum_{b\in\Gamma} p(a,b)\right)\log\left(\sum_{c\in\Gamma} p(a,c)\right) = \mathrm{H}(\mathsf{X}), \end{aligned} \tag{5.37}$$

as required. □

Remark 5.12. It should be noted that Proposition 5.11 does not carry over to the von Neumann entropy of quantum states (cf. Theorem 5.27).

The next theorem represents a direct and straightforward application of Lemma 5.4. A quantum analogue of this theorem, which is stated and proved in Section 5.2.3, is not known to have nearly so straightforward a proof.

Theorem 5.13. *Let Σ be an alphabet and let $u_0, u_1, v_0, v_1 \in [0,\infty)^\Sigma$ be vectors of nonnegative real numbers indexed by Σ. It holds that*

$$\mathrm{D}(u_0 + u_1 \| v_0 + v_1) \leq \mathrm{D}(u_0\|v_0) + \mathrm{D}(u_1\|v_1). \tag{5.38}$$

Proof. By Lemma 5.4 it holds that

$$\begin{aligned} &\mathrm{D}(u_0 + u_1 \| v_0 + v_1) \\ &\quad = \frac{1}{\ln(2)} \sum_{a\in\Sigma} \theta(u_0(a) + u_1(a), v_0(a) + v_1(a)) \\ &\quad \leq \frac{1}{\ln(2)} \sum_{a\in\Sigma} \big(\theta(u_0(a), v_0(a)) + \theta(u_1(a), v_1(a))\big) \\ &\quad = \mathrm{D}(u_0\|v_0) + \mathrm{D}(u_1\|v_1), \end{aligned} \tag{5.39}$$

as claimed. □

For all vectors $u, v \in [0,\infty)^\Sigma$ and scalars $\alpha, \beta \in [0,\infty)$ it holds that

$$\mathrm{D}(\alpha u \| \beta v) = \alpha\,\mathrm{D}(u\|v) + \frac{1}{\ln(2)}\theta(\alpha,\beta)\sum_{a\in\Sigma} u(a), \tag{5.40}$$

provided one makes the interpretation $0 \cdot \infty = 0$ in the case that $\alpha = 0$ and $\mathrm{D}(u\|v) = \infty$, or in the case that $\theta(\alpha, \beta) = \infty$ and $u = 0$. This can be verified through a direct calculation. As $\theta(\alpha, \alpha) = 0$ for all $\alpha \in [0, \infty)$, one obtains the identity

$$\mathrm{D}(\alpha u \| \alpha v) = \alpha \, \mathrm{D}(u\|v), \tag{5.41}$$

where again it is to be interpreted that $0 \cdot \infty = 0$. Alternately, one may verify that this identity holds by observing

$$\theta(\alpha\beta, \alpha\gamma) = \alpha \, \theta(\beta, \gamma) \tag{5.42}$$

for all nonnegative real numbers $\alpha, \beta, \gamma \in [0, \infty)$. Through this identity, one obtains the following corollary to Theorem 5.13.

Corollary 5.14 (Joint convexity of the relative entropy)**.** *Let Σ be an alphabet, let $u_0, u_1, v_0, v_1 \in [0, \infty)^\Sigma$ be vectors of nonnegative real numbers indexed by Σ, and let $\lambda \in [0, 1]$. It holds that*

$$\begin{aligned} &\mathrm{D}(\lambda u_0 + (1-\lambda) u_1 \| \lambda v_0 + (1-\lambda) v_1) \\ &\qquad \leq \lambda \, \mathrm{D}(u_0\|v_0) + (1-\lambda) \, \mathrm{D}(u_1\|v_1). \end{aligned} \tag{5.43}$$

Through a similar argument, one may prove that the relative entropy of one vector with respect to another cannot increase under the action of any stochastic operation performed simultaneously on the two vectors.

Theorem 5.15. *Let Σ and Γ be alphabets, let $u, v \in [0, \infty)^\Sigma$ be vectors, and let $A \in \mathrm{L}\big(\mathbb{R}^\Sigma, \mathbb{R}^\Gamma\big)$ be a stochastic operator. It holds that*

$$\mathrm{D}(Au\|Av) \leq \mathrm{D}(u\|v). \tag{5.44}$$

Proof. By Lemma 5.4 along with the identity (5.42), it holds that

$$\begin{aligned} \mathrm{D}(Au\|Av) &= \frac{1}{\ln(2)} \sum_{a \in \Gamma} \theta\left(\sum_{b \in \Sigma} A(a,b) u(b), \sum_{b \in \Sigma} A(a,b) v(b) \right) \\ &\leq \frac{1}{\ln(2)} \sum_{a \in \Gamma} \sum_{b \in \Sigma} A(a,b) \, \theta(u(b), v(b)) \\ &= \frac{1}{\ln(2)} \sum_{b \in \Sigma} \theta(u(b), v(b)) \\ &= \mathrm{D}(u\|v), \end{aligned} \tag{5.45}$$

as required. □

Quantitative bounds on Shannon entropy and relative entropy

Two bounds, one concerning the Shannon entropy and one concerning the relative entropy, will now be proved. The first bound is a quantitative form of the statement that the Shannon entropy function is continuous on the set of all probability vectors.

Theorem 5.16 (Audenaert)**.** *Let $p_0, p_1 \in \mathcal{P}(\Sigma)$ be probability vectors, for Σ being an alphabet with $|\Sigma| \geq 2$. It holds that*

$$|\mathrm{H}(p_0) - \mathrm{H}(p_1)| \leq \lambda \log(|\Sigma| - 1) + \mathrm{H}(\lambda, 1 - \lambda) \tag{5.46}$$

for $\lambda = \frac{1}{2} \| p_0 - p_1 \|_1$.

Proof. The theorem holds trivially when $p_0 = p_1$, so it will be assumed that this is not the case. Let $\Sigma_0, \Sigma_1 \subseteq \Sigma$ be disjoint sets defined as

$$\begin{aligned} \Sigma_0 &= \big\{a \in \Sigma \,:\, p_0(a) > p_1(a)\big\}, \\ \Sigma_1 &= \big\{a \in \Sigma \,:\, p_0(a) < p_1(a)\big\}, \end{aligned} \tag{5.47}$$

and let vectors $u_0, u_1 \in [0,1]^{\Sigma}$ be defined as

$$u_0(a) = \begin{cases} p_0(a) - p_1(a) & \text{if } a \in \Sigma_0 \\ 0 & \text{otherwise,} \end{cases} \tag{5.48}$$

$$u_1(a) = \begin{cases} p_1(a) - p_0(a) & \text{if } a \in \Sigma_1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.49}$$

for every $a \in \Sigma$. It holds that $p_0 - p_1 = u_0 - u_1$ and $u_0(a)u_1(a) = 0$ for all $a \in \Sigma$, and moreover

$$\sum_{a \in \Sigma} u_0(a) = \lambda = \sum_{a \in \Sigma} u_1(a). \tag{5.50}$$

Taking $w \in [0,1]^{\Sigma}$ to be defined as

$$w(a) = \min\{p_0(a), p_1(a)\} \tag{5.51}$$

for every $a \in \Sigma$, one finds that $p_0 = u_0 + w$, $p_1 = u_1 + w$, and

$$\sum_{a \in \Sigma} w(a) = 1 - \lambda. \tag{5.52}$$

Next, observe that the identity

$$\begin{aligned}(\alpha+\beta)\log(\alpha+\beta) - \alpha\log(\alpha) - \beta\log(\beta)& \\ = (\alpha+\beta)\,\mathrm{H}\biggl(\frac{\alpha}{\alpha+\beta}, \frac{\beta}{\alpha+\beta}\biggr)&\end{aligned} \tag{5.53}$$

holds for every choice of nonnegative real numbers α and β, assuming at least one of them is positive (and, as is to be expected, interpreting $0\log(0)$ as 0 if either α or β is 0). Through this identity, the following two expressions are obtained:

$$\mathrm{H}(u_0) + \mathrm{H}(w) - \mathrm{H}(p_0) = \sum_{a\in\Sigma_0} p_0(a)\,\mathrm{H}\biggl(\frac{u_0(a)}{p_0(a)}, \frac{w(a)}{p_0(a)}\biggr), \tag{5.54}$$

$$\mathrm{H}(u_1) + \mathrm{H}(w) - \mathrm{H}(p_1) = \sum_{a\in\Sigma_1} p_1(a)\,\mathrm{H}\biggl(\frac{u_1(a)}{p_1(a)}, \frac{w(a)}{p_1(a)}\biggr). \tag{5.55}$$

In both cases, the restriction of the sums to the sets Σ_0 and Σ_1 reflects the exclusion of 0 summands. Both sums include only nonnegative summands, and therefore

$$\mathrm{H}(p_0) \leq \mathrm{H}(u_0) + \mathrm{H}(w) \quad \text{and} \quad \mathrm{H}(p_1) \leq \mathrm{H}(u_1) + \mathrm{H}(w). \tag{5.56}$$

Furthermore, by setting

$$\alpha_0 = \sum_{a\in\Sigma_0} p_0(a) \quad \text{and} \quad \alpha_1 = \sum_{a\in\Sigma_1} p_1(a), \tag{5.57}$$

one has that $\alpha_0, \alpha_1 \in [\lambda, 1]$, and the following two inequalities are obtained from the concavity of the Shannon entropy (Proposition 5.5):

$$\mathrm{H}(u_0) + \mathrm{H}(w) - \mathrm{H}(p_0) \leq \alpha_0\,\mathrm{H}\biggl(\frac{\lambda}{\alpha_0}, 1 - \frac{\lambda}{\alpha_0}\biggr), \tag{5.58}$$

$$\mathrm{H}(u_1) + \mathrm{H}(w) - \mathrm{H}(p_1) \leq \alpha_1\,\mathrm{H}\biggl(\frac{\lambda}{\alpha_1}, 1 - \frac{\lambda}{\alpha_1}\biggr). \tag{5.59}$$

Given that the function

$$f_\lambda(\alpha) = \alpha\,\mathrm{H}\biggl(\frac{\lambda}{\alpha}, 1 - \frac{\lambda}{\alpha}\biggr) \tag{5.60}$$

is strictly increasing on the interval $[\lambda, 1]$, it follows that

$$\begin{aligned}0 \leq \mathrm{H}(u_0) + \mathrm{H}(w) - \mathrm{H}(p_0) \leq \mathrm{H}(\lambda, 1-\lambda),& \\ 0 \leq \mathrm{H}(u_1) + \mathrm{H}(w) - \mathrm{H}(p_1) \leq \mathrm{H}(\lambda, 1-\lambda).&\end{aligned} \tag{5.61}$$

By the triangle inequality together with (5.61), one may therefore conclude that

$$\begin{aligned} &|\mathrm{H}(p_0) - \mathrm{H}(p_1)| - |\mathrm{H}(u_0) - \mathrm{H}(u_1)| \\ &\quad \leq |(\mathrm{H}(p_0) - \mathrm{H}(u_0) - \mathrm{H}(w)) - (\mathrm{H}(p_1) - \mathrm{H}(u_1) - \mathrm{H}(w))| \\ &\quad \leq \mathrm{H}(\lambda, 1-\lambda). \end{aligned} \tag{5.62}$$

To complete the proof, it suffices to prove

$$|\mathrm{H}(u_0) - \mathrm{H}(u_1)| \leq \lambda \log(|\Sigma| - 1). \tag{5.63}$$

For any alphabet Γ and any vector $v \in [0,\infty)^\Gamma$ with

$$\sum_{b \in \Gamma} v(b) = \lambda, \tag{5.64}$$

it holds that

$$-\lambda \log(\lambda) \leq \mathrm{H}(v) \leq \lambda \log(|\Gamma|) - \lambda \log(\lambda), \tag{5.65}$$

as was demonstrated in Proposition 5.9. Given that u_0 and u_1 are supported on disjoint subsets of Σ and have entries summing to the same value λ, it follows that

$$|\mathrm{H}(u_0) - \mathrm{H}(u_1)| \leq \lambda \log(|\Gamma|) - \lambda \log(\lambda) + \lambda \log(\lambda) = \lambda \log(|\Gamma|), \tag{5.66}$$

for Γ being a proper subset of Σ. The largest value obtained for the upper bound occurs when Γ has one fewer element than Σ, yielding the required inequality (5.63), which completes the proof. □

The second bound, which concerns the relative entropy function, is a quantitative form of Proposition 5.7. It lower-bounds the relative entropy $\mathrm{D}(p_0 \| p_1)$, for probability vectors p_0 and p_1, by a quantity determined by their 1-norm distance $\| p_0 - p_1 \|_1$.

Theorem 5.17 (Pinsker's inequality)**.** *Let Σ be an alphabet and $p_0, p_1 \in \mathcal{P}(\Sigma)$ be probability vectors indexed by Σ. It holds that*

$$\mathrm{D}(p_0 \| p_1) \geq \frac{1}{2\ln(2)} \| p_0 - p_1 \|_1^2. \tag{5.67}$$

The proof of Theorem 5.17 will make use of the following lemma, which is equivalent to a special case of the theorem in which $|\Sigma| = 2$.

Lemma 5.18. *For all choices of real numbers $\alpha, \beta \in [0,1]$ it holds that*

$$\theta(\alpha,\beta) + \theta(1-\alpha,1-\beta) \geq 2(\alpha-\beta)^2. \tag{5.68}$$

Proof. The inequality in the statement of the lemma is immediate in the case that $\beta \in \{0,1\}$. In the case that $\alpha \in \{0,1\}$ and $\beta \in (0,1)$, the inequality in the statement of the lemma is equivalent to

$$-\ln(\beta) \geq 2(1-\beta)^2, \tag{5.69}$$

which can be verified using elementary calculus. It remains to consider the case where $\alpha, \beta \in (0,1)$. Under this assumption it may be verified that

$$\begin{aligned} &\theta(\alpha,\beta) + \theta(1-\alpha,1-\beta) \\ &\qquad = (\eta(\beta) + \eta(1-\beta)) - (\eta(\alpha) + \eta(1-\alpha)) \\ &\qquad\quad + (\alpha-\beta)(\eta'(\beta) - \eta'(1-\beta)) \\ &\qquad = f(\beta) - f(\alpha) + (\alpha-\beta)f'(\beta) \end{aligned} \tag{5.70}$$

for $f : [0,1] \to \mathbb{R}$ defined as $f(\gamma) = \eta(\gamma) + \eta(1-\gamma)$ for all $\gamma \in [0,1]$. By Taylor's theorem it holds that

$$f(\alpha) = f(\beta) + (\alpha-\beta)f'(\beta) + \frac{1}{2}(\alpha-\beta)^2 f''(\gamma) \tag{5.71}$$

for some choice of γ being a convex combination of α and β. Equation (5.71) therefore holds for some choice of $\gamma \in (0,1)$. Evaluating the second derivative of f yields

$$f''(\gamma) = -\left(\frac{1}{\gamma} + \frac{1}{1-\gamma}\right), \tag{5.72}$$

whereby it follows that $f''(\gamma) \leq -4$ for all $\gamma \in (0,1)$. This implies the inequality (5.68), which completes the proof. □

Proof of Theorem 5.17. Define disjoint sets $\Sigma_0, \Sigma_1, \Gamma \subseteq \Sigma$ as

$$\Sigma_0 = \{a \in \Sigma : p_0(a) > p_1(a)\}, \tag{5.73}$$

$$\Sigma_1 = \{a \in \Sigma : p_0(a) < p_1(a)\}, \tag{5.74}$$

$$\Gamma = \{a \in \Sigma : p_0(a) = p_1(a)\}, \tag{5.75}$$

and define a stochastic operator $A \in \mathrm{L}\big(\mathbb{R}^{\{0,1\}}, \mathbb{R}^{\Sigma}\big)$ as

$$A = \sum_{a\in\Sigma_0} E_{0,a} + \sum_{a\in\Sigma_1} E_{1,a} + \frac{1}{2}\sum_{a\in\Gamma}(E_{0,a} + E_{1,a}). \tag{5.76}$$

Let $\alpha = (Ap_0)(0)$ and $\beta = (Ap_1)(0)$, and note that $(Ap_0)(1) = 1 - \alpha$ and $(Ap_1)(1) = 1 - \beta$, as p_0 and p_1 are probability vectors and A is stochastic. It holds that

$$\alpha - \beta = \sum_{a \in \Sigma_0} (p_0(a) - p_1(a)) = \sum_{a \in \Sigma_1} (p_1(a) - p_0(a)) = \frac{1}{2} \| p_0 - p_1 \|_1. \quad (5.77)$$

By Theorem 5.15 and Lemma 5.18, one finds that

$$\begin{aligned} \mathrm{D}(p_0 \| p_1) \geq \mathrm{D}(Ap_0 \| Ap_1) &= \frac{1}{\ln(2)} \left(\theta(\alpha, \beta) + \theta(1 - \alpha, 1 - \beta) \right) \\ &\geq \frac{2}{\ln(2)} (\alpha - \beta)^2 = \frac{1}{2 \ln(2)} \| p_0 - p_1 \|_1^2, \end{aligned} \quad (5.78)$$

as required. □

5.2 Quantum entropy

The von Neumann entropy and quantum relative entropy functions, which extend the Shannon entropy and relative entropy functions from nonnegative vectors to positive semidefinite operators, are defined in this section. Fundamental properties of these functions are established, including the key properties of joint convexity of the quantum relative entropy and strong subadditivity of the von Neumann entropy.

5.2.1 Definitions of quantum entropic functions

The von Neumann entropy function represents a natural extension of the Shannon entropy function from nonnegative vectors to positive semidefinite operators; as the following definition states, the von Neumann entropy is defined as the Shannon entropy of a given positive semidefinite operator's vector of eigenvalues.

Definition 5.19. Let $\mathcal{X}$ be a complex Euclidean space and let $P \in \mathrm{Pos}(\mathcal{X})$ be a positive semidefinite operator. The *von Neumann entropy* of P is defined as

$$\mathrm{H}(P) = \mathrm{H}(\lambda(P)), \quad (5.79)$$

for $\lambda(P)$ being the vector of eigenvalues of P.

The von Neumann entropy may also be expressed as

$$\mathrm{H}(P) = -\operatorname{Tr}\left(P \log(P)\right). \tag{5.80}$$

Formally speaking, this expression assumes that the operator $P\log(P)$ is defined for all positive semidefinite operators $P \in \operatorname{Pos}(\mathcal{X})$, despite the fact that $\log(P)$ is only defined for positive definite operators P. The natural interpretation is that $P\log(P)$ refers to the operator obtained by extending the scalar function

$$\alpha \mapsto \begin{cases} \alpha \log(\alpha) & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha = 0 \end{cases} \tag{5.81}$$

to positive semidefinite operators in the usual way (q.v. Section 1.1.3).

Similar to the Shannon entropy usually being considered for probability vectors, it is most common that one considers the von Neumann entropy function on density operator inputs. Also similar to the Shannon entropy, it is convenient to speak of the von Neumann entropy $\mathrm{H}(\mathsf{X})$ of a register X, which means the quantity $\mathrm{H}(\rho)$ for $\rho \in \mathrm{D}(\mathcal{X})$ representing the state of X at the moment being considered. Once again, the notation $\mathrm{H}(\mathsf{X},\mathsf{Y})$ is taken to mean $\mathrm{H}((\mathsf{X},\mathsf{Y}))$, and likewise for other forms of compound registers.

The study of the von Neumann entropy is aided by the consideration of the *quantum relative entropy*, which is an extension of the ordinary relative entropy from vectors to positive semidefinite operators.

Definition 5.20. Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \operatorname{Pos}(\mathcal{X})$ be positive semidefinite operators. The *quantum relative entropy* of P with respect to Q, which is denoted $\mathrm{D}(P\|Q)$, is defined as follows:

$$\mathrm{D}(P\|Q) = \begin{cases} \operatorname{Tr}(P\log(P)) - \operatorname{Tr}(P\log(Q)) & \text{if } \operatorname{im}(P) \subseteq \operatorname{im}(Q) \\ \infty & \text{otherwise.} \end{cases} \tag{5.82}$$

This definition is deserving of a short explanation because, as before, the logarithm is really only defined for positive definite operators. However, the operator $P\log(Q)$ has a natural interpretation for positive semidefinite operators P and Q that satisfy $\operatorname{im}(P) \subseteq \operatorname{im}(Q)$. The action of this operator on the subspace $\operatorname{im}(Q)$ is well-defined, as Q is a positive definite operator when restricted to this subspace, while its action on the subspace $\ker(Q)$ is taken to be the zero operator. This interpretation is equivalent to identifying $0\log(0)$ with 0, as the condition $\operatorname{im}(P) \subseteq \operatorname{im}(Q)$ implies that P acts as the

zero mapping on $\ker(Q)$. The operator $P\log(P)$ is defined for all positive semidefinite operators P, as was discussed previously.

It will be convenient to make note of a concrete expression for the value $\mathrm{D}(P\|Q)$, assuming $\mathrm{im}(P) \subseteq \mathrm{im}(Q)$. Let $n = \dim(\mathcal{X})$ and suppose that

$$P = \sum_{j=1}^{n} \lambda_j(P)\, x_j x_j^* \quad \text{and} \quad Q = \sum_{k=1}^{n} \lambda_k(Q)\, y_k y_k^* \tag{5.83}$$

are spectral decompositions of P and Q. Let $r = \mathrm{rank}(P)$ and $s = \mathrm{rank}(Q)$, and observe that the expressions of P and Q in (5.83) may be truncated to r and s terms, respectively. It then holds that

$$\mathrm{D}(P\|Q) = \sum_{j=1}^{r} \sum_{k=1}^{s} |\langle x_j, y_k\rangle|^2\, \lambda_j(P)\big(\log(\lambda_j(P)) - \log(\lambda_k(Q))\big). \tag{5.84}$$

The omission of the indices $j \in \{r+1,\ldots,n\}$ and $k \in \{s+1,\ldots,n\}$ in the sums is consistent with the identification $0\log(0) = 0$ suggested above. In particular, if k is such that $\lambda_k(Q) = 0$, then it must hold that

$$|\langle x_j, y_k\rangle|^2 \lambda_j(P) = 0 \tag{5.85}$$

for all $j \in \{1,\ldots,n\}$ by the assumption $\mathrm{im}(P) \subseteq \mathrm{im}(Q)$. An alternative expression for the quantum relative entropy $\mathrm{D}(P\|Q)$, for P and Q having spectral decompositions (5.83), which is valid for all choices of P and Q, is given by

$$\mathrm{D}(P\|Q) = \frac{1}{\ln(2)} \sum_{j=1}^{n} \sum_{k=1}^{n} \theta\Big(|\langle x_j, y_k\rangle|^2 \lambda_j(P), |\langle x_j, y_k\rangle|^2 \lambda_k(Q)\Big). \tag{5.86}$$

The *conditional von Neumann entropy* and *quantum mutual information* are defined in an analogous manner to the conditional Shannon entropy and mutual information. More precisely, for two registers X and Y in a given state of interest, one defines the conditional von Neumann entropy of X given Y as

$$\mathrm{H}(\mathsf{X}|\mathsf{Y}) = \mathrm{H}(\mathsf{X},\mathsf{Y}) - \mathrm{H}(\mathsf{Y}), \tag{5.87}$$

and one defines the quantum mutual information between X and Y as

$$\mathrm{I}(\mathsf{X}:\mathsf{Y}) = \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}). \tag{5.88}$$

5.2.2 Elementary properties of quantum entropic functions

This section discusses elementary properties of the von Neumann entropy and quantum relative entropy functions. Specifically, these are properties that may be established without making essential use of the joint convexity of the quantum relative entropy, which is proved in the section following this one, or other equivalent statements.

Continuity of the von Neumann entropy

The von Neumann entropy function is continuous, owing to the fact that it is a composition of continuous functions: the Shannon entropy function is continuous at every point in its domain, as is the function

$$\lambda : \mathrm{Herm}(\mathcal{X}) \to \mathbb{R}^n, \tag{5.89}$$

for $n = \dim(\mathcal{X})$.

Simple identities concerning quantum entropy

The three propositions that follow are stated as propositions for the sake of convenience. They may be verified directly through the definitions of the von Neumann entropy and quantum relative entropy functions.

Proposition 5.21. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces for which it holds that $\dim(\mathcal{X}) \leq \dim(\mathcal{Y})$, let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, and let $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y})$ be an isometry. It holds that*

$$\mathrm{H}\big(VPV^*\big) = \mathrm{H}(P) \quad \textit{and} \quad \mathrm{D}\big(VPV^* \big\| VQV^*\big) = \mathrm{D}(P\|Q). \tag{5.90}$$

Proposition 5.22. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $P \in \mathrm{Pos}(\mathcal{X})$ and $Q \in \mathrm{Pos}(\mathcal{Y})$ be positive semidefinite operators. It holds that*

$$\mathrm{H}\left(\begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}\right) = \mathrm{H}(P) + \mathrm{H}(Q) \tag{5.91}$$

and

$$\mathrm{H}(P \otimes Q) = \mathrm{Tr}(Q)\,\mathrm{H}(P) + \mathrm{Tr}(P)\,\mathrm{H}(Q). \tag{5.92}$$

In particular, it holds that

$$\mathrm{H}(\rho \otimes \sigma) = \mathrm{H}(\rho) + \mathrm{H}(\sigma) \tag{5.93}$$

for all choices of density operators $\rho \in \mathrm{D}(\mathcal{X})$ and $\sigma \in \mathrm{D}(\mathcal{Y})$.

Proposition 5.23. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $P_0, Q_0 \in \mathrm{Pos}(\mathcal{X})$ and $P_1, Q_1 \in \mathrm{Pos}(\mathcal{Y})$ be positive semidefinite operators, and assume that P_0 and P_1 are nonzero. It holds that*

$$\mathrm{D}(P_0 \otimes P_1 \| Q_0 \otimes Q_1) = \mathrm{Tr}(P_1)\, \mathrm{D}(P_0 \| Q_0) + \mathrm{Tr}(P_0)\, \mathrm{D}(P_1 \| Q_1). \tag{5.94}$$

As a consequence of the tensor product identities in the second and third of these propositions, one finds that the following two identities hold for all choices of a complex Euclidean space $\mathcal{X}$, positive semidefinite operators $P, Q \in \mathrm{Pos}(\mathcal{X})$, and scalars $\alpha, \beta \in (0, \infty)$:

$$\mathrm{H}(\alpha P) = \alpha\, \mathrm{H}(P) - \alpha \log(\alpha)\, \mathrm{Tr}(P), \tag{5.95}$$

$$\mathrm{D}(\alpha P \| \beta Q) = \alpha\, \mathrm{D}(P \| Q) + \alpha \log(\alpha / \beta)\, \mathrm{Tr}(P). \tag{5.96}$$

Klein's inequality

An analogous statement to Proposition 5.7 in the quantum setting is known as *Klein's inequality*. It implies that the quantum relative entropy function is nonnegative for density operator inputs.

Proposition 5.24 (Klein's inequality)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators satisfying $\mathrm{Tr}(P) \geq \mathrm{Tr}(Q)$. It holds that $\mathrm{D}(P \| Q) \geq 0$. In particular, it holds that $\mathrm{D}(\rho \| \sigma) \geq 0$ for every choice of density operators $\rho, \sigma \in \mathrm{D}(\mathcal{X})$.*

Proof. Let $n = \dim(\mathcal{X})$ and let

$$P = \sum_{j=1}^{n} \lambda_j(P)\, x_j x_j^* \quad \text{and} \quad Q = \sum_{k=1}^{n} \lambda_k(Q)\, y_k y_k^* \tag{5.97}$$

be spectral decompositions of P and Q. By Lemma 5.4, it holds that

$$\begin{aligned} \mathrm{D}(P \| Q) &= \frac{1}{\ln(2)} \sum_{j,k} \theta\Big(|\langle x_j, y_k \rangle|^2 \lambda_j(P),\, |\langle x_j, y_k \rangle|^2 \lambda_k(Q) \Big) \\ &\geq \frac{1}{\ln(2)}\, \theta\Bigg(\sum_{j,k} |\langle x_j, y_k \rangle|^2\, \lambda_j(P), \sum_{j,k} |\langle x_j, y_k \rangle|^2\, \lambda_k(Q) \Bigg) \\ &= \frac{1}{\ln(2)}\, \theta(\mathrm{Tr}(P), \mathrm{Tr}(Q)), \end{aligned} \tag{5.98}$$

where the sums are over all $j, k \in \{1, \ldots, n\}$. If it holds that $\mathrm{Tr}(P) \geq \mathrm{Tr}(Q)$ then $\theta(\mathrm{Tr}(P), \mathrm{Tr}(Q)) \geq 0$, which completes the proof. □

Concavity and subadditivity of von Neumann entropy

Similar to the Shannon entropy, the von Neumann entropy is concave and subadditive, as the following two theorems establish.

Theorem 5.25 (Concavity of von Neumann entropy)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, and let $\lambda \in [0,1]$. It holds that*

$$\mathrm{H}(\lambda P + (1-\lambda)Q) \geq \lambda \,\mathrm{H}(P) + (1-\lambda)\,\mathrm{H}(Q). \tag{5.99}$$

Proof. A straightforward computation reveals that

$$\mathrm{D}\left(\begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \middle\| \begin{pmatrix} \frac{P+Q}{2} & 0 \\ 0 & \frac{P+Q}{2} \end{pmatrix}\right) = 2\,\mathrm{H}\left(\frac{P+Q}{2}\right) - \mathrm{H}(P) - \mathrm{H}(Q). \tag{5.100}$$

As the operators

$$\begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{P+Q}{2} & 0 \\ 0 & \frac{P+Q}{2} \end{pmatrix} \tag{5.101}$$

have the same trace, the quantity represented by (5.100) is nonnegative by Klein's inequality (Proposition 5.24). It therefore holds that

$$\mathrm{H}\left(\frac{P+Q}{2}\right) \geq \frac{1}{2}\,\mathrm{H}(P) + \frac{1}{2}\,\mathrm{H}(Q) \tag{5.102}$$

which implies that the von Neumann entropy is midpoint concave on the domain $\mathrm{Pos}(\mathcal{X})$. As the von Neumann entropy function is continuous on all of $\mathrm{Pos}(\mathcal{X})$, it follows that it is in fact a concave function on this domain, which completes the proof. □

Theorem 5.26 (Subadditivity of von Neumann entropy)**.** *Let X and Y be registers. For every state of the register (X,Y), it holds that*

$$\mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}). \tag{5.103}$$

Proof. The inequality in the statement of the proposition may equivalently be written

$$\mathrm{H}(\rho) \leq \mathrm{H}(\rho[\mathsf{X}]) + \mathrm{H}(\rho[\mathsf{Y}]) \tag{5.104}$$

for $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ denoting an arbitrary state of the pair (X,Y). Using the formula

$$\log(P \otimes Q) = \log(P) \otimes \mathbb{1} + \mathbb{1} \otimes \log(Q), \tag{5.105}$$

together with the fact that $\operatorname{im}(\rho) \subseteq \operatorname{im}(\rho[\mathsf{X}] \otimes \rho[\mathsf{Y}])$, it may be observed that

$$\mathrm{D}\big(\rho \big\| \rho[\mathsf{X}] \otimes \rho[\mathsf{Y}]\big) = -\mathrm{H}(\rho) + \mathrm{H}(\rho[\mathsf{X}]) + \mathrm{H}(\rho[\mathsf{Y}]). \tag{5.106}$$

It holds that (5.106) is nonnegative by Klein's inequality (Proposition 5.24), and therefore the inequality (5.104) follows. □

Von Neumann entropy and purifications

Let X and Y be registers, and assume the compound register (X, Y) is in a pure state uu^*, for $u \in \mathcal{X} \otimes \mathcal{Y}$ being a unit vector. By means of the Schmidt decomposition, one may write

$$u = \sum_{a \in \Sigma} \sqrt{p(a)}\, x_a \otimes y_a \tag{5.107}$$

for some choice of an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and orthonormal sets

$$\{x_a : a \in \Sigma\} \subset \mathcal{X} \quad \text{and} \quad \{y_a : a \in \Sigma\} \subset \mathcal{Y}. \tag{5.108}$$

It holds that

$$(uu^*)[\mathsf{X}] = \sum_{a \in \Sigma} p(a)\, x_a x_a^* \quad \text{and} \quad (uu^*)[\mathsf{Y}] = \sum_{a \in \Sigma} p(a)\, y_a y_a^*, \tag{5.109}$$

and therefore

$$\mathrm{H}(\mathsf{X}) = \mathrm{H}(p) = \mathrm{H}(\mathsf{Y}). \tag{5.110}$$

This simple observation, when combined with the notion of purifications of states, provides a useful tool for reasoning about the von Neumann entropy of collections of registers. The proof of the following theorem offers one example along these lines.

Theorem 5.27. *Let X and Y be registers. For every state of the register (X, Y), it holds that*

$$\mathrm{H}(\mathsf{X}) \leq \mathrm{H}(\mathsf{Y}) + \mathrm{H}(\mathsf{X}, \mathsf{Y}). \tag{5.111}$$

Proof. Let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be a state of the pair (X, Y), and introduce a new register Z whose associated complex Euclidean space $\mathcal{Z}$ has dimension at least $\operatorname{rank}(\rho)$. By Theorem 2.9, there must exist a unit vector $u \in \mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}$ such that

$$\rho = \mathrm{Tr}_{\mathcal{Z}}(uu^*). \tag{5.112}$$

Now, consider the situation in which the compound register $(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$ is in the pure state uu^*, which is consistent with the state of (X, Y) being ρ by the requirement (5.112). By the argument suggested above, one finds that

$$\mathrm{H}(\mathsf{X}) = \mathrm{H}(\mathsf{Y}, \mathsf{Z}) \quad \text{and} \quad \mathrm{H}(\mathsf{X}, \mathsf{Y}) = \mathrm{H}(\mathsf{Z}). \tag{5.113}$$

By the subadditivity of the von Neumann entropy (Theorem 5.26), one has

$$\mathrm{H}(\mathsf{Y}, \mathsf{Z}) \leq \mathrm{H}(\mathsf{Y}) + \mathrm{H}(\mathsf{Z}), \tag{5.114}$$

and therefore

$$\mathrm{H}(\mathsf{X}) \leq \mathrm{H}(\mathsf{Y}) + \mathrm{H}(\mathsf{X}, \mathsf{Y}). \tag{5.115}$$

The required inequality has therefore been established for all choices of the state ρ, which completes the proof. □

The Fannes–Audenaert inequality

The next theorem establishes an upper bound on the difference between the values of the von Neumann entropy function of two density operators. It may be seen as a quantitative form of the statement that the von Neumann entropy is continuous, restricted to density operator inputs. It is essentially a quantum generalization of Theorem 5.16, and its proof is based on that theorem.

Theorem 5.28 (Fannes–Audenaert inequality)**.** *Let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be density operators, for $\mathcal{X}$ being a complex Euclidean space of dimension $n \geq 2$, and let*

$$\delta = \frac{1}{2} \left\| \rho_0 - \rho_1 \right\|_1. \tag{5.116}$$

It holds that

$$|\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| \leq \delta \log(n-1) + \mathrm{H}(\delta, 1-\delta). \tag{5.117}$$

The following lemma relating the trace distance between two Hermitian operators to the 1-norm distance between vectors of their eigenvalues is used to reduce Theorem 5.28 to Theorem 5.16.

Lemma 5.29. *Let $X, Y \in \mathrm{Herm}(\mathcal{X})$ be Hermitian operators, for $\mathcal{X}$ being a complex Euclidean space of dimension n. It holds that*

$$\sum_{k=1}^{n} |\lambda_k(X) - \lambda_k(Y)| \leq \left\| X - Y \right\|_1 \leq \sum_{k=1}^{n} |\lambda_k(X) - \lambda_{n-k+1}(Y)|. \tag{5.118}$$

Proof. To prove the first inequality, let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be operators providing a Jordan–Hahn decomposition $X - Y = P - Q$, and let $Z = P + Y$ (which is equivalent to $Z = Q + X$). As $Z \geq X$, it follows from the Courant–Fischer theorem (Theorem 1.2) that $\lambda_k(Z) \geq \lambda_k(X)$ for all $k \in \{1, \ldots, n\}$. Thus,

$$\begin{aligned} \lambda_k(X) - \lambda_k(Y) &\leq \big(\lambda_k(X) - \lambda_k(Y)\big) + 2\big(\lambda_k(Z) - \lambda_k(X)\big) \\ &= 2\lambda_k(Z) - \big(\lambda_k(X) + \lambda_k(Y)\big). \end{aligned} \tag{5.119}$$

By similar reasoning it follows that

$$\lambda_k(Y) - \lambda_k(X) \leq 2\lambda_k(Z) - \big(\lambda_k(X) + \lambda_k(Y)\big), \tag{5.120}$$

and therefore

$$|\lambda_k(X) - \lambda_k(Y)| \leq 2\lambda_k(Z) - \big(\lambda_k(X) + \lambda_k(Y)\big). \tag{5.121}$$

Consequently, one has

$$\begin{aligned} \sum_{k=1}^{n} |\lambda_k(X) - \lambda_k(Y)| &\leq \sum_{k=1}^{n} \big(2\lambda_k(Z) - \big(\lambda_k(X) + \lambda_k(Y)\big)\big) \\ &= 2\,\mathrm{Tr}(Z) - \mathrm{Tr}(X) - \mathrm{Tr}(Y) = \mathrm{Tr}(P) + \mathrm{Tr}(Q) = \big\|X - Y\big\|_1. \end{aligned} \tag{5.122}$$

To prove the second inequality, observe that

$$\|X - Y\|_1 = \langle 2\Pi - \mathbb{1}, X - Y \rangle \tag{5.123}$$

for some choice of a projection operator Π, owing to the fact that $X - Y$ is Hermitian. Let $r = \mathrm{rank}(\Pi)$, and note the following two inequalities:

$$\begin{aligned} \langle \Pi, X \rangle &\leq \lambda_1(X) + \cdots + \lambda_r(X), \\ \langle \Pi, Y \rangle &\geq \lambda_{n-r+1}(Y) + \cdots + \lambda_n(Y). \end{aligned} \tag{5.124}$$

It follows that

$$\begin{aligned} &\|X - Y\|_1 \\ &\quad \leq 2\big(\lambda_1(X) + \cdots + \lambda_r(X)\big) - 2\big(\lambda_{n-r+1}(Y) + \cdots + \lambda_n(Y)\big) \\ &\qquad\quad - \mathrm{Tr}(X) + \mathrm{Tr}(Y) \\ &\quad = \sum_{k=1}^{r} (\lambda_k(X) - \lambda_{n-k+1}(Y)) + \sum_{k=r+1}^{n} (\lambda_{n-k+1}(Y) - \lambda_k(X)) \\ &\quad \leq \sum_{k=1}^{n} |\lambda_k(X) - \lambda_{n-k+1}(Y)|, \end{aligned} \tag{5.125}$$

as required. □

Proof of Theorem 5.28. Define $\delta_0, \delta_1 \in [0,1]$ as

$$\delta_0 = \sum_{k=1}^{n} |\lambda_k(\rho_0) - \lambda_k(\rho_1)| \quad \text{and} \quad \delta_1 = \sum_{k=1}^{n} |\lambda_k(\rho_0) - \lambda_{n-k+1}(\rho_1)|. \tag{5.126}$$

By Lemma 5.29 it holds that $\delta_0 \leq \delta \leq \delta_1$, and therefore $\delta = \alpha\delta_0 + (1-\alpha)\delta_1$ for some choice of $\alpha \in [0,1]$. By Theorem 5.16 it holds that

$$\begin{aligned} &|\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| \\ &\qquad = |\mathrm{H}(\lambda_1(\rho_0), \ldots, \lambda_n(\rho_0)) - \mathrm{H}(\lambda_1(\rho_1), \ldots, \lambda_n(\rho_1))| \\ &\qquad \leq \delta_1 \log(n-1) + \mathrm{H}(\delta_1, 1-\delta_1) \end{aligned} \tag{5.127}$$

and

$$\begin{aligned} &|\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| \\ &\qquad = |\mathrm{H}(\lambda_1(\rho_0), \ldots, \lambda_n(\rho_0)) - \mathrm{H}(\lambda_n(\rho_1), \ldots, \lambda_1(\rho_1))| \\ &\qquad \leq \delta_0 \log(n-1) + \mathrm{H}(\delta_0, 1-\delta_0). \end{aligned} \tag{5.128}$$

Thus, by the concavity of the Shannon entropy function (Proposition 5.5), it follows that

$$\begin{aligned} |\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| &\leq (\alpha\delta_0 + (1-\alpha)\delta_1)\log(n-1) \\ &\qquad + \alpha\,\mathrm{H}(\delta_0, 1-\delta_0) + (1-\alpha)\,\mathrm{H}(\delta_1, 1-\delta_1) \\ &\leq \delta \log(n-1) + \mathrm{H}(\delta, 1-\delta), \end{aligned} \tag{5.129}$$

as required. □

The Fannes–Audenaert inequality is saturated for all values of $\delta \in [0,1]$ and $n \geq 2$. For instance, for any choice of $n \geq 2$ and $\Sigma = \{1, \ldots, n\}$, one may consider the density operators

$$\rho_0 = E_{1,1} \quad \text{and} \quad \rho_1 = (1-\delta)E_{1,1} + \frac{\delta}{n-1}\sum_{k=2}^{n} E_{k,k}. \tag{5.130}$$

It holds that

$$\delta = \frac{1}{2}\|\rho_0 - \rho_1\|_1 \tag{5.131}$$

and

$$|\mathrm{H}(\rho_0) - \mathrm{H}(\rho_1)| = \mathrm{H}(\rho_1) = \mathrm{H}(\delta, 1-\delta) + \delta\log(n-1), \tag{5.132}$$

which saturates the Fannes–Audenaert inequality.

The quantum relative entropy as a limit of difference quotients

As the following proposition states, the quantum relative entropy can be expressed as the limit of a simple expression of its arguments. This fact will be useful in Section 5.2.3, for the task of proving that the quantum relative entropy is jointly convex.

Proposition 5.30. *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{D}(P\|Q) = \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P) - \langle P^{1-\varepsilon}, Q^{\varepsilon} \rangle}{\varepsilon}. \tag{5.133}$$

Proof. The proposition is immediate in the case that $\mathrm{im}(P) \not\subseteq \mathrm{im}(Q)$, for in this case

$$\lim_{\varepsilon \downarrow 0} \Big(\mathrm{Tr}(P) - \langle P^{1-\varepsilon}, Q^{\varepsilon} \rangle \Big) = \langle P, \mathbb{1} - \Pi_{\mathrm{im}(Q)} \rangle \tag{5.134}$$

is a positive real number. This implies that the limit in (5.133) evaluates to positive infinity, which is in agreement with the quantum relative entropy. The proposition is also immediate in the case that $P = 0$. It therefore remains to consider the case that P is a nonzero operator and $\mathrm{im}(P) \subseteq \mathrm{im}(Q)$, which is taken as an assumption for the remainder of the proof.

Let $r = \mathrm{rank}(P)$ and $s = \mathrm{rank}(Q)$. By the spectral theorem (as stated by Corollary 1.4), one may write

$$P = \sum_{j=1}^{r} \lambda_j(P)\, x_j x_j^* \quad \text{and} \quad Q = \sum_{k=1}^{s} \lambda_k(Q)\, y_k y_k^* \tag{5.135}$$

for orthonormal collections of vectors $\{x_1, \ldots, x_r\}$ and $\{y_1, \ldots, y_s\}$. Define a function $f : \mathbb{R} \to \mathbb{R}$ as

$$f(\alpha) = \sum_{j=1}^{r} \sum_{k=1}^{s} |\langle x_j, y_k \rangle|^2 \, \lambda_j(P)^{1-\alpha} \, \lambda_k(Q)^{\alpha} \tag{5.136}$$

for all $\alpha \in \mathbb{R}$. This function is differentiable at every point $\alpha \in \mathbb{R}$, with its derivative given by

$$f'(\alpha) = -\sum_{j=1}^{r} \sum_{k=1}^{s} |\langle x_j, y_k \rangle|^2 \, \lambda_j(P)^{1-\alpha} \, \lambda_k(Q)^{\alpha} \ln\left(\frac{\lambda_j(P)}{\lambda_k(Q)} \right). \tag{5.137}$$

Now, it holds that

$$f(\alpha) = \left\langle P^{1-\alpha}, Q^{\alpha} \right\rangle \tag{5.138}$$

for every $\alpha \in (0,1)$, while

$$f(0) = \left\langle P, \Pi_{\mathrm{im}(Q)} \right\rangle = \mathrm{Tr}(P). \tag{5.139}$$

Evaluating the derivative of f at 0 yields

$$f'(0) = -\ln(2)\, \mathrm{D}(P\|Q), \tag{5.140}$$

while the definition of the derivative, as the limit of difference quotients, yields

$$f'(0) = \lim_{\varepsilon \downarrow 0} \frac{f(\varepsilon) - f(0)}{\varepsilon} = \lim_{\varepsilon \downarrow 0} \frac{\langle P^{1-\varepsilon}, Q^{\varepsilon} \rangle - \mathrm{Tr}(P)}{\varepsilon}. \tag{5.141}$$

The proposition follows by combining equations (5.141) and (5.140). □

5.2.3 Joint convexity of quantum relative entropy

This section contains a proof of a fundamental fact concerning the quantum relative entropy, which is that it is a jointly convex function. By making use of this key fact, one may prove that several other important properties of the von Neumann entropy and quantum relative entropy functions hold.

Proof of the joint convexity of the quantum relative entropy

Multiple proofs of the joint convexity of the quantum relative entropy are known. The proof to be presented below will make use of the following technical lemma relating the diagonal and off-diagonal blocks of any 2-by-2 positive semidefinite block operator, under the assumption that the blocks are Hermitian and the diagonal blocks commute.

Lemma 5.31. *Let $\mathcal{X}$ be a complex Euclidean space, let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators such that $[P, Q] = 0$, and let $H \in \mathrm{Herm}(\mathcal{X})$ be a Hermitian operator for which*

$$\begin{pmatrix} P & H \\ H & Q \end{pmatrix} \in \mathrm{Pos}(\mathcal{X} \oplus \mathcal{X}). \tag{5.142}$$

It holds that $H \leq \sqrt{P}\sqrt{Q}$.

Proof. The lemma will first be proved for P and Q being positive definite operators. By Lemma 3.18 it follows that

$$\left\| P^{-\frac{1}{2}} H Q^{-\frac{1}{2}} \right\| \leq 1, \tag{5.143}$$

which implies that every eigenvalue of the operator $P^{-\frac{1}{2}} H Q^{-\frac{1}{2}}$ is bounded by 1 in absolute value. As P and Q commute, it holds that the eigenvalues of $P^{-\frac{1}{4}} Q^{-\frac{1}{4}} H Q^{-\frac{1}{4}} P^{-\frac{1}{4}}$ agree with those of $P^{-\frac{1}{2}} H Q^{-\frac{1}{2}}$, and therefore

$$\lambda_1 \left(P^{-\frac{1}{4}} Q^{-\frac{1}{4}} H Q^{-\frac{1}{4}} P^{-\frac{1}{4}} \right) \leq 1. \tag{5.144}$$

The inequality (5.144) is equivalent to

$$P^{-\frac{1}{4}} Q^{-\frac{1}{4}} H Q^{-\frac{1}{4}} P^{-\frac{1}{4}} \leq \mathbb{1}, \tag{5.145}$$

which, again by the commutativity of P and Q, implies $H \leq \sqrt{P}\sqrt{Q}$.

In the general case where P and Q are not necessarily positive definite, the argument above may be applied to $P + \varepsilon \mathbb{1}$ and $Q + \varepsilon \mathbb{1}$ in place of P and Q, respectively, to obtain

$$H \leq \sqrt{P + \varepsilon \mathbb{1}} \sqrt{Q + \varepsilon \mathbb{1}} \tag{5.146}$$

for all $\varepsilon > 0$. The function $\varepsilon \mapsto \sqrt{P + \varepsilon \mathbb{1}} \sqrt{Q + \varepsilon \mathbb{1}} - H$ is continuous on the domain $[0, \infty)$, and so the preimage of the closed set $\mathrm{Pos}(\mathcal{X})$ under this function is closed. Given that every $\varepsilon > 0$ is contained in this preimage, it follows that 0 is contained in the preimage as well: $\sqrt{P}\sqrt{Q} - H$ is positive semidefinite, which proves the lemma. □

The next step toward the joint convexity of the quantum relative entropy is to prove the following theorem. It is one formulation of a fact known as *Lieb's concavity theorem*.

Theorem 5.32 (Lieb's concavity theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $A_0, A_1 \in \mathrm{Pos}(\mathcal{X})$ and $B_0, B_1 \in \mathrm{Pos}(\mathcal{Y})$ be positive semidefinite operators. For every choice of a real number $\alpha \in [0, 1]$ it holds that*

$$(A_0 + A_1)^{\alpha} \otimes (B_0 + B_1)^{1-\alpha} \geq A_0^{\alpha} \otimes B_0^{1-\alpha} + A_1^{\alpha} \otimes B_1^{1-\alpha}. \tag{5.147}$$

Remark 5.33. Within the context of this theorem, as well as its proof, one should make the interpretation $P^0 = \Pi_{\mathrm{im}(P)}$ for every positive semidefinite operator P.

Proof of Theorem 5.32. For every real number $\alpha \in [0,1]$, define operators as follows:

$$\begin{aligned} X(\alpha) &= A_0^{\alpha} \otimes B_0^{1-\alpha}, \\ Y(\alpha) &= A_1^{\alpha} \otimes B_1^{1-\alpha}, \\ Z(\alpha) &= (A_0 + A_1)^{\alpha} \otimes (B_0 + B_1)^{1-\alpha}. \end{aligned} \tag{5.148}$$

The operators within these three individual collections commute, meaning

$$[X(\alpha), X(\beta)] = 0, \quad [Y(\alpha), Y(\beta)] = 0, \quad \text{and} \quad [Z(\alpha), Z(\beta)] = 0 \tag{5.149}$$

for every choice of $\alpha, \beta \in [0,1]$, and moreover it holds that

$$\sqrt{X(\alpha)}\sqrt{X(\beta)} = X\left(\frac{\alpha+\beta}{2}\right), \tag{5.150}$$

$$\sqrt{Y(\alpha)}\sqrt{Y(\beta)} = Y\left(\frac{\alpha+\beta}{2}\right), \tag{5.151}$$

$$\sqrt{Z(\alpha)}\sqrt{Z(\beta)} = Z\left(\frac{\alpha+\beta}{2}\right). \tag{5.152}$$

With respect to these operators, the statement of the theorem is equivalent to the claim that

$$Z(\alpha) \geq X(\alpha) + Y(\alpha) \tag{5.153}$$

for every $\alpha \in [0,1]$. The function

$$\alpha \mapsto Z(\alpha) - (X(\alpha) + Y(\alpha)) \tag{5.154}$$

defined on the interval $[0,1]$ is continuous, and therefore the preimage of the closed set $\mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ under this function is closed. It therefore suffices to prove that the set of all $\alpha \in [0,1]$ for which (5.153) holds is dense in $[0,1]$.

Now, suppose it has been proved that

$$Z(\alpha) \geq X(\alpha) + Y(\alpha) \quad \text{and} \quad Z(\beta) \geq X(\beta) + Y(\beta) \tag{5.155}$$

for some particular choice of real numbers $\alpha, \beta \in [0,1]$. It holds that

$$\begin{pmatrix} \sqrt{X(\alpha)} \\ \sqrt{X(\beta)} \end{pmatrix} \begin{pmatrix} \sqrt{X(\alpha)} & \sqrt{X(\beta)} \end{pmatrix} = \begin{pmatrix} X(\alpha) & X\left(\frac{\alpha+\beta}{2}\right) \\ X\left(\frac{\alpha+\beta}{2}\right) & X(\beta) \end{pmatrix} \tag{5.156}$$

is positive semidefinite, and likewise

$$\begin{pmatrix} \sqrt{Y(\alpha)} \\ \sqrt{Y(\beta)} \end{pmatrix} \begin{pmatrix} \sqrt{Y(\alpha)} & \sqrt{Y(\beta)} \end{pmatrix} = \begin{pmatrix} Y(\alpha) & Y\left(\frac{\alpha+\beta}{2}\right) \\ Y\left(\frac{\alpha+\beta}{2}\right) & Y(\beta) \end{pmatrix} \tag{5.157}$$

is positive semidefinite. The sum of these two matrices is therefore positive semidefinite, and given the inequalities (5.155) it therefore follows that

$$\begin{pmatrix} Z(\alpha) & X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) \\ X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) & Z(\beta) \end{pmatrix} \tag{5.158}$$

is positive semidefinite. Invoking Lemma 5.31, one finds that

$$X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right) \leq \sqrt{Z(\alpha)}\sqrt{Z(\beta)} = Z\left(\frac{\alpha+\beta}{2}\right). \tag{5.159}$$

It trivially holds that $Z(0) \geq X(0) + Y(0)$ and $Z(1) \geq X(1) + Y(1)$. For any choice of $\alpha, \beta \in [0,1]$, one has that the inequalities (5.155) together imply that

$$Z\left(\frac{\alpha+\beta}{2}\right) \geq X\left(\frac{\alpha+\beta}{2}\right) + Y\left(\frac{\alpha+\beta}{2}\right). \tag{5.160}$$

The inequality (5.153) must therefore hold for every $\alpha \in [0,1]$ taking the form $\alpha = k/2^n$ for nonnegative integers k and n with $k \leq 2^n$. The set of all such α is dense in $[0,1]$, so the theorem is proved. □

Corollary 5.34. *Let $P_0, P_1, Q_0, Q_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, for $\mathcal{X}$ being any complex Euclidean space. It holds that*

$$\left\langle (P_0 + P_1)^\alpha, (Q_0 + Q_1)^{1-\alpha} \right\rangle \geq \left\langle P_0^\alpha, Q_0^{1-\alpha} \right\rangle + \left\langle P_1^\alpha, Q_1^{1-\alpha} \right\rangle \tag{5.161}$$

for every $\alpha \in [0,1]$.

Proof. By making the substituting $A_0 = P_0$, $A_1 = P_1$, $B_0 = Q_0^{\mathsf{T}}$, and $B_1 = Q_1^{\mathsf{T}}$ in Theorem 5.32, one finds that

$$(P_0 + P_1)^\alpha \otimes (Q_0^{\mathsf{T}} + Q_1^{\mathsf{T}})^{1-\alpha} \geq P_0^\alpha \otimes (Q_0^{\mathsf{T}})^{1-\alpha} + P_1^\alpha \otimes (Q_1^{\mathsf{T}})^{1-\alpha}, \tag{5.162}$$

and therefore

$$\begin{aligned} &\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \left((P_0 + P_1)^\alpha \otimes (Q_0^{\mathsf{T}} + Q_1^{\mathsf{T}})^{1-\alpha} \right) \mathrm{vec}(\mathbb{1}_{\mathcal{X}}) \\ &\quad \geq \mathrm{vec}(\mathbb{1}_{\mathcal{X}})^* \left(P_0^\alpha \otimes (Q_0^{\mathsf{T}})^{1-\alpha} + P_1^\alpha \otimes (Q_1^{\mathsf{T}})^{1-\alpha} \right) \mathrm{vec}(\mathbb{1}_{\mathcal{X}}). \end{aligned} \tag{5.163}$$

Simplifying the two sides of this inequality yields (5.161), as required. □

The joint convexity of the quantum relative entropy now follows from a combination of Corollary 5.34 with Proposition 5.30.

Theorem 5.35. *Let $P_0, P_1, Q_0, Q_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, for $\mathcal{X}$ being any complex Euclidean space. It holds that*

$$\mathrm{D}(P_0 + P_1 \| Q_0 + Q_1) \leq \mathrm{D}(P_0 \| Q_0) + \mathrm{D}(P_1 \| Q_1). \tag{5.164}$$

Proof. By Proposition 5.30 together with Corollary 5.34 it holds that

$$\begin{aligned}
&\mathrm{D}(P_0 + P_1 \| Q_0 + Q_1) \\
&\quad = \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P_0 + P_1) - \langle (P_0 + P_1)^{1-\varepsilon}, (Q_0 + Q_1)^{\varepsilon} \rangle}{\varepsilon} \\
&\quad \leq \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P_0 + P_1) - \langle P_0^{1-\varepsilon}, Q_0^{\varepsilon} \rangle - \langle P_1^{1-\varepsilon}, Q_1^{\varepsilon} \rangle}{\varepsilon} \\
&\quad = \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P_0) - \langle P_0^{1-\varepsilon}, Q_0^{\varepsilon} \rangle}{\varepsilon} + \frac{1}{\ln(2)} \lim_{\varepsilon \downarrow 0} \frac{\mathrm{Tr}(P_1) - \langle P_1^{1-\varepsilon}, Q_1^{\varepsilon} \rangle}{\varepsilon} \\
&\quad = \mathrm{D}(P_0 \| Q_0) + \mathrm{D}(P_1 \| Q_1),
\end{aligned} \tag{5.165}$$

which proves the theorem. □

Corollary 5.36 (Joint convexity of quantum relative entropy)**.** *Let $\mathcal{X}$ be a complex Euclidean space, let P_0, P_1, Q_0, $Q_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, and let $\lambda \in [0, 1]$. It holds that*

$$\begin{aligned}
&\mathrm{D}(\lambda P_0 + (1-\lambda) P_1 \| \lambda Q_0 + (1-\lambda) Q_1) \\
&\qquad \leq \lambda \, \mathrm{D}(P_0 \| Q_0) + (1-\lambda) \, \mathrm{D}(P_1 \| Q_1).
\end{aligned} \tag{5.166}$$

Proof. Combining Theorem 5.35 with the identity (5.96) yields

$$\begin{aligned}
&\mathrm{D}(\lambda P_0 + (1-\lambda) P_1 \| \lambda Q_0 + (1-\lambda) Q_1) \\
&\qquad \leq \mathrm{D}(\lambda P_0 \| \lambda Q_0) + \mathrm{D}((1-\lambda) P_1 \| (1-\lambda) Q_1) \\
&\qquad = \lambda \, \mathrm{D}(P_0 \| Q_0) + (1-\lambda) \, \mathrm{D}(P_1 \| Q_1),
\end{aligned} \tag{5.167}$$

as required. □

Monotonicity of quantum relative entropy

As was suggested above, the fact that the quantum relative entropy function is jointly convex has several interesting implications. One such implication is that the quantum relative entropy function is monotonically decreasing under the action of any channel. The next proposition establishes that this is so for mixed-unitary channels, and the theorem that follows establishes that the same is true for all channels.

Proposition 5.37. *Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{C}(\mathcal{X})$ be a mixed-unitary channel, and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{D}\big(\Phi(P)\big\|\Phi(Q)\big) \leq \mathrm{D}(P\|Q). \tag{5.168}$$

Proof. As Φ is a mixed-unitary channel, there must exist an alphabet Σ, a collection of unitary operators $\{U_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{X})$, and a probability vector $p \in \mathcal{P}(\Sigma)$, such that

$$\Phi(X) = \sum_{a\in\Sigma} p(a) U_a X U_a^* \tag{5.169}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Applying Corollary 5.36, along with Proposition 5.21, one has

$$\begin{aligned}
\mathrm{D}(\Phi(P)\|\Phi(Q)) &= \mathrm{D}\Bigg(\sum_{a\in\Sigma} p(a) U_a P U_a^* \,\Bigg\|\, \sum_{a\in\Sigma} p(a) U_a Q U_a^*\Bigg) \\
&\leq \sum_{a\in\Sigma} p(a)\, \mathrm{D}\big(U_a P U_a^* \big\| U_a Q U_a^*\big) \\
&= \sum_{a\in\Sigma} p(a)\, \mathrm{D}(P\|Q) \\
&= \mathrm{D}(P\|Q),
\end{aligned} \tag{5.170}$$

as required. □

Theorem 5.38 (Monotonicity of quantum relative entropy)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\mathrm{D}(\Phi(P)\|\Phi(Q)) \leq \mathrm{D}(P\|Q). \tag{5.171}$$

Proof. By Corollary 2.27 there must exist a complex Euclidean space $\mathcal{Z}$ and a linear isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ for which

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(AXA^*\big) \tag{5.172}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Let $\Omega \in \mathrm{C}(\mathcal{Z})$ denote the completely depolarizing channel, defined by

$$\Omega(Z) = \mathrm{Tr}(Z)\omega \tag{5.173}$$

for all $Z \in \mathrm{L}(\mathcal{Z})$, where

$$\omega = \frac{\mathbb{1}_{\mathcal{Z}}}{\dim(\mathcal{Z})} \tag{5.174}$$

denotes the completely mixed state with respect to the space $\mathcal{Z}$. As was demonstrated in Section 4.1.1, the channel Ω is a mixed-unitary channel, from which it follows that $\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Omega$ is also a mixed-unitary channel. By Proposition 5.37, together with Proposition 5.21, it therefore holds that

$$\begin{aligned} &\mathrm{D}\big((\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Omega)(APA^*)\big\|(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Omega)(AQA^*)\big) \\ &\qquad \leq \mathrm{D}(APA^*\|AQA^*) = \mathrm{D}(P\|Q). \end{aligned} \tag{5.175}$$

As

$$(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Omega)(AXA^*) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \otimes \omega = \Phi(X) \otimes \omega \tag{5.176}$$

for all $X \in \mathrm{L}(\mathcal{X})$, it follows by Proposition 5.23 that

$$\mathrm{D}(\Phi(P)\|\Phi(Q)) = \mathrm{D}\big(\Phi(P) \otimes \omega \,\big\|\, \Phi(Q) \otimes \omega\big) \leq \mathrm{D}(P\|Q), \tag{5.177}$$

which completes the proof. □

Strong subadditivity of von Neumann entropy

Another implication of the joint convexity of quantum relative entropy is the following theorem, stating that the von Neumann entropy possesses a property known as *strong subadditivity*.

Theorem 5.39 (Strong subadditivity of von Neumann entropy)**.** *Let* X, Y, *and* Z *be registers. For every state of the register* $(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$ *it holds that*

$$\mathrm{H}(\mathsf{X}, \mathsf{Y}, \mathsf{Z}) + \mathrm{H}(\mathsf{Z}) \leq \mathrm{H}(\mathsf{X}, \mathsf{Z}) + \mathrm{H}(\mathsf{Y}, \mathsf{Z}). \tag{5.178}$$

Proof. Let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ be chosen arbitrarily and let

$$\omega = \frac{\mathbb{1}_{\mathcal{X}}}{\dim(\mathcal{X})} \tag{5.179}$$

denote the completely mixed state with respect to the space $\mathcal{X}$. These two equalities hold:

$$\begin{aligned} &\mathrm{D}\big(\rho[\mathsf{X},\mathsf{Y},\mathsf{Z}] \,\big\|\, \omega \otimes \rho[\mathsf{Y},\mathsf{Z}]\big) \\ &\qquad = -\mathrm{H}\big(\rho[\mathsf{X},\mathsf{Y},\mathsf{Z}]\big) + \mathrm{H}\big(\rho[\mathsf{Y},\mathsf{Z}]\big) + \log(\dim(\mathcal{X})) \end{aligned} \tag{5.180}$$

and

$$\begin{aligned} &\mathrm{D}\big(\rho[\mathsf{X},\mathsf{Z}] \,\big\|\, \omega \otimes \rho[\mathsf{Z}]\big) \\ &\qquad = -\mathrm{H}\big(\rho[\mathsf{X},\mathsf{Z}]\big) + \mathrm{H}\big(\rho[\mathsf{Z}]\big) + \log(\dim(\mathcal{X})). \end{aligned} \tag{5.181}$$

Taking the channel $\Phi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z}, \mathcal{X} \otimes \mathcal{Z})$ to be the partial trace over $\mathcal{Y}$ in Theorem 5.38, one finds that

$$\mathrm{D}\big(\rho[\mathsf{X},\mathsf{Z}] \,\big\|\, \omega \otimes \rho[\mathsf{Z}]\big) \leq \mathrm{D}\big(\rho[\mathsf{X},\mathsf{Y},\mathsf{Z}] \,\big\|\, \omega \otimes \rho[\mathsf{Y},\mathsf{Z}]\big), \tag{5.182}$$

and therefore

$$\mathrm{H}(\rho[\mathsf{X},\mathsf{Y},\mathsf{Z}]) + \mathrm{H}(\rho[\mathsf{Z}]) \leq \mathrm{H}(\rho[\mathsf{X},\mathsf{Z}]) + \mathrm{H}(\rho[\mathsf{Y},\mathsf{Z}]), \tag{5.183}$$

which proves the theorem. □

The corollary that follows gives an equivalent statement to the strong subadditivity of von Neumann entropy, stated in terms of the quantum mutual information.

Corollary 5.40. *Let* X, Y, *and* Z *be registers. For every state of the register* $(\mathsf{X},\mathsf{Y},\mathsf{Z})$ *it holds that*

$$\mathrm{I}(\mathsf{X}:\mathsf{Y}) \leq \mathrm{I}(\mathsf{X}:\mathsf{Y},\mathsf{Z}). \tag{5.184}$$

Proof. By Theorem 5.39 it holds that

$$\mathrm{H}(\mathsf{X},\mathsf{Y},\mathsf{Z}) + \mathrm{H}(\mathsf{Y}) \leq \mathrm{H}(\mathsf{X},\mathsf{Y}) + \mathrm{H}(\mathsf{Y},\mathsf{Z}), \tag{5.185}$$

which is equivalent to

$$\mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{Y},\mathsf{Z}) - \mathrm{H}(\mathsf{X},\mathsf{Y},\mathsf{Z}). \tag{5.186}$$

Adding $\mathrm{H}(\mathsf{X})$ to both sides gives

$$\mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}) \leq \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y},\mathsf{Z}) - \mathrm{H}(\mathsf{X},\mathsf{Y},\mathsf{Z}). \tag{5.187}$$

This inequality is equivalent to (5.184), which completes the proof. □

The quantum Pinsker inequality

The final implication of the joint convexity of quantum relative entropy to be presented in this section is a quantum analogue of Theorem 5.17 that establishes a lower bound on the quantum relative entropy between two density operators in terms of their trace distance.

Theorem 5.41 (Quantum Pinsker inequality)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$ be density operators. It holds that*

$$\mathrm{D}(\rho_0 \| \rho_1) \geq \frac{1}{2\ln(2)} \left\| \rho_0 - \rho_1 \right\|_1^2. \tag{5.188}$$

Proof. Let $\Sigma = \{0,1\}$, let $P_0, P_1 \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators providing the Jordan–Hahn decomposition $\rho_0 - \rho_1 = P_0 - P_1$, and define a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$ as

$$\mu(0) = \Pi_{\mathrm{im}(P_0)} \quad \text{and} \quad \mu(1) = \mathbb{1} - \Pi_{\mathrm{im}(P_0)}. \tag{5.189}$$

This measurement is optimal for discriminating between the states ρ_0 and ρ_1 given with equal probability, as discussed in Section 3.1.1. For probability vectors $p_0, p_1 \in \mathcal{P}(\Sigma)$ defined as

$$p_0(a) = \langle \mu(a), \rho_0 \rangle \quad \text{and} \quad p_1(a) = \langle \mu(a), \rho_1 \rangle \tag{5.190}$$

for each $a \in \Sigma$, one has

$$\left\| p_0 - p_1 \right\|_1 = \left\| \rho_0 - \rho_1 \right\|_1. \tag{5.191}$$

Now let $\Phi \in \mathrm{C}(\mathcal{X}, \mathbb{C}^\Sigma)$ be the quantum-to-classical channel associated with μ, which satisfies

$$\Phi(X) = \langle \mu(0), X \rangle E_{0,0} + \langle \mu(1), X \rangle E_{1,1} \tag{5.192}$$

for each $X \in \mathrm{L}(\mathcal{X})$. It holds that

$$\mathrm{D}(\rho_0 \| \rho_1) \geq \mathrm{D}(\Phi(\rho_0) \| \Phi(\rho_1)) \tag{5.193}$$

by Corollary 5.36. By Theorem 5.17 it holds that

$$\begin{aligned} \mathrm{D}(\Phi(\rho_0) \| \Phi(\rho_1)) &= \mathrm{D}(\mathrm{Diag}(p_0) \| \mathrm{Diag}(p_1)) = \mathrm{D}(p_0 \| p_1) \\ &\geq \frac{1}{2\ln(2)} \left\| p_0 - p_1 \right\|_1^2 = \frac{1}{2\ln(2)} \left\| \rho_0 - \rho_1 \right\|_1^2, \end{aligned} \tag{5.194}$$

which completes the proof. □

5.3 Source coding

This section discusses the notion of *source coding*, as it relates to quantum information, and to the von Neumann entropy function in particular. The term *source coding*, as it is interpreted here, refers to the process of encoding information produced by given source in such a way that it may later be decoded. One natural goal of such a process is to compress the information produced by the source, in order to reduce costs of storage or transmission. Three principal variants of source coding will be discussed.

The first is a purely classical variant in which information from a given classical source is encoded into a fixed-length binary string in such a way that the information produced by the source can be decoded with high probability. A theorem known as *Shannon's source coding theorem* establishes asymptotic bounds on compression rates that are achievable for this task, given a standard assumption on the source.

The second variant of source coding to be discussed is a quantum analogue to the first; a source produces quantum information that is to be encoded into a sequence of qubits and then decoded. A theorem due to Schumacher, representing a quantum analogue of Shannon's source coding theorem, establishes asymptotic bounds on the rates of compression that are achievable for this task.

The third variant of source coding to be considered is one in which a source produces classical information, which is encoded into the quantum state of a collection of registers, and then decoded through a measurement performed on these registers. Theorems due to Holevo and Nayak establish fundamental limitations on two specific formulations of this task.

5.3.1 Classical source coding

In the first variant of source coding to be considered in the present section, a classical source produces a sequence of symbols, chosen independently from a known probability distribution. This sequence is to be encoded into a binary string in such a way that it may later be decoded, revealing the original sequence produced by the source with high probability.

The main purpose of this discussion, as it pertains to this book, is to introduce basic concepts and techniques regarding classical source coding that will carry over to the analogous quantum variant of this task. With this

purpose in mind, the discussion is limited to *fixed-length* coding schemes. These are schemes in which the length of each encoding is determined only by the number of symbols produced by the source, and not by the symbols themselves. A typical goal when designing such a scheme is to minimize the length of the binary string encodings while allowing for a recovery of the original sequence with high probability.

Shannon's source coding theorem establishes a fundamental connection between the rates of compression that can be achieved by such schemes and the Shannon entropy of the probability vector describing the source. While Shannon's source coding theorem is often stated in terms of *variable-length* coding schemes, with which one aims for a perfect recovery of the symbols produced by the source while minimizing the expected length of the binary string encodings, the fixed-length variant presented below translates more directly to the quantum setting.

Coding schemes and the statement of Shannon's source coding theorem

Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let $\Gamma = \{0,1\}$ denote the binary alphabet. For any choice of a positive integer n and real numbers $\alpha > 0$ and $\delta \in (0,1)$, and for $m = \lfloor \alpha n \rfloor$, a pair of mappings

$$\begin{aligned} f &: \Sigma^n \to \Gamma^m \\ g &: \Gamma^m \to \Sigma^n \end{aligned} \tag{5.195}$$

is said to be an (n, α, δ)-*coding scheme* for p if and only if it holds that

$$\sum_{a_1 \cdots a_n \in G} p(a_1) \cdots p(a_n) > 1 - \delta, \tag{5.196}$$

for

$$G = \big\{ a_1 \cdots a_n \in \Sigma^n \,:\, g(f(a_1 \cdots a_n)) = a_1 \cdots a_n \big\}. \tag{5.197}$$

(Here, and throughout the remainder of this chapter, elements of sets of the form Σ^n are written as strings $a_1 \cdots a_n$ rather than n-tuples $(a_1, \ldots, a_n)$, and likewise for Cartesian products of other alphabets.)

The expression on the left-hand side of (5.196) represents the probability that a random choice of symbols $a_1, \ldots, a_n \in \Sigma$, with each symbol chosen independently according to the probability vector p, results in a sequence satisfying

$$g(f(a_1 \cdots a_n)) = a_1 \cdots a_n. \tag{5.198}$$

The following scenario describes an abstract setting in which such coding schemes may be considered.

Scenario 5.42. Alice has a device (the source) that sequentially generates symbols chosen at random from an alphabet Σ. Each randomly generated symbol is independently distributed according to a single probability vector p. Alice allows the device to produce a string of n symbols $a_1 \cdots a_n$, and aims to communicate this string to Bob using as few bits of communication as possible.

To do this, Alice and Bob will use a coding scheme taking the form (5.195), which is assumed to have been agreed upon before the random generation of the symbols $a_1 \cdots a_n$. Alice *encodes* $a_1 \cdots a_n$ into a string of $m = \lfloor \alpha n \rfloor$ bits by computing $f(a_1 \cdots a_n)$, and sends the resulting binary string $f(a_1 \cdots a_n)$ to Bob. Bob *decodes* the string by applying the function g, obtaining $g(f(a_1 \cdots a_n))$. The coding scheme is said to be *correct* in the event that (5.198) holds, which is equivalent to $a_1 \cdots a_n \in G$, for then Bob will have obtained the correct string $a_1 \cdots a_n$.

If it is the case that the pair (f, g) is an (n, α, δ)-coding scheme for p, then the number δ is an upper bound on the probability that the coding scheme fails to be correct, so that Bob does not recover the string Alice obtained from the source, while α represents the average number of bits (as the value of n increases) needed to encode each symbol.

For a given probability vector p, it is evident that an (n, α, δ)-coding scheme will exist for some choices of the parameters n, α, and δ, and not others. The range of values of α for which coding schemes exist is closely related to the Shannon entropy $\mathrm{H}(p)$, as the following theorem establishes.

Theorem 5.43 (Shannon's source coding theorem)**.** *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let $\alpha > 0$ and $\delta \in (0, 1)$ be real numbers. The following statements hold:*

1. *If $\alpha > \mathrm{H}(p)$, then there exists an (n, α, δ)-coding scheme for p for all but finitely many choices of $n \in \mathbb{N}$.*
2. *If $\alpha < \mathrm{H}(p)$, then there exists an (n, α, δ)-coding scheme for p for at most finitely many choices of $n \in \mathbb{N}$.*

A proof of this theorem is presented below, following a discussion of the notion of a *typical string*, which is central to the proof. The general notion of

typicality, which can be formalized in various specific ways, will also play a major role in Chapter 8, which is devoted to the topic of quantum channel capacities.

Typical strings

The notion of a typical string, for a given distribution of symbols, a string length, and an error parameter, is defined as follows.

Definition 5.44. Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\varepsilon > 0$ be a positive real number. A string $a_1 \cdots a_n \in \Sigma^n$ is said to be *ϵ-typical (with respect to p)* if and only if

$$2^{-n(\mathrm{H}(p)+\varepsilon)} < p(a_1) \cdots p(a_n) < 2^{-n(\mathrm{H}(p)-\varepsilon)}. \tag{5.199}$$

The notation $T_{n,\varepsilon}(p)$ refers to the set of all strings $a_1 \cdots a_n \in \Sigma^n$ for which the inequalities (5.199) hold, and when the probability vector p can safely be taken as being implicit, one may write $T_{n,\varepsilon}$ rather than $T_{n,\varepsilon}(p)$.

A random selection of a string $a_1 \cdots a_n \in \Sigma^n$, with each symbol being independently distributed according to $p \in \mathcal{P}(\Sigma)$, is increasingly likely to be ε-typical as n grows, as the following proposition demonstrates.

Proposition 5.45. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let $\varepsilon > 0$. It holds that*

$$\lim_{n\to\infty} \sum_{a_1\cdots a_n \in T_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) = 1. \tag{5.200}$$

Proof. Let $Y_1, \ldots, Y_n$ be independent and identically distributed random variables, defined as follows: one first chooses $a \in \Sigma$ randomly according to the probability vector p, and then sets the value of the random variable to be the real number $-\log(p(a))$ for whichever value of a was selected. It holds that the expected value of each Y_k is

$$\mathrm{E}(Y_k) = -\sum_{a\in\Sigma} p(a)\log(p(a)) = \mathrm{H}(p). \tag{5.201}$$

The conclusion of the lemma may now be written

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n} Y_k - \mathrm{H}(p)\right| \geq \varepsilon\right) = 0, \tag{5.202}$$

which is true by the weak law of large numbers (Theorem 1.19). □

The proposition that follows establishes an upper bound on the number of ε-typical strings of a given length.

Proposition 5.46. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let $\varepsilon > 0$ be a positive real number, and let n be a positive integer. It holds that*

$$\bigl|T_{n,\varepsilon}(p)\bigr| < 2^{n(\mathrm{H}(p)+\varepsilon)}. \tag{5.203}$$

Proof. By the definition of ε-typicality, one has

$$1 \geq \sum_{a_1\cdots a_n \in T_{n,\varepsilon}(p)} p(a_1)\cdots p(a_n) > 2^{-n(\mathrm{H}(p)+\varepsilon)}\,|T_{n,\varepsilon}(p)|, \tag{5.204}$$

and therefore $|T_{n,\varepsilon}(p)| < 2^{n(\mathrm{H}(p)+\varepsilon)}$. □

Proof of Shannon's source coding theorem

Shannon's source coding theorem (Theorem 5.43) can be proved through a conceptually argument—a suitable coding scheme may be obtained for sufficiently large values of n by assigning a unique binary string to each typical string, with every other string being encoded arbitrarily. Conversely, any coding scheme that fails to account for a large fraction of the typical strings can be shown to fail with high probability.

Proof of Theorem 5.43. Assume first that $\alpha > \mathrm{H}(p)$, and choose $\varepsilon > 0$ so that $\alpha > \mathrm{H}(p) + 2\varepsilon$. A coding scheme of the form

$$\begin{aligned} f_n &: \Sigma^n \to \Gamma^m \\ g_n &: \Gamma^m \to \Sigma^n, \end{aligned} \tag{5.205}$$

for $m = \lfloor \alpha n \rfloor$, will be defined for every $n \in \mathbb{N}$ satisfying $n > 1/\varepsilon$. Observe, for each $n > 1/\varepsilon$, that the assumption $\alpha > \mathrm{H}(p) + 2\varepsilon$ implies that

$$m = \lfloor \alpha n \rfloor > n(\mathrm{H}(p) + \varepsilon). \tag{5.206}$$

By Proposition 5.46 it holds that

$$|T_{n,\varepsilon}| < 2^{n(\mathrm{H}(p)+\varepsilon)} < 2^m, \tag{5.207}$$

and one may therefore define a function $f_n : \Sigma^n \to \Gamma^m$ that is injective when restricted to $T_{n,\varepsilon}$, together with a function $g_n : \Gamma^m \to \Sigma^n$ that is chosen so that $g_n(f_n(a_1\cdots a_n)) = a_1\cdots a_n$ for every $a_1\cdots a_n \in T_{n,\varepsilon}$. Thus, for

$$G_n = \{a_1\cdots a_n \in \Sigma^n \,:\, g_n(f_n(a_1\cdots a_n)) = a_1\cdots a_n\}, \tag{5.208}$$

it holds that $T_{n,\varepsilon} \subseteq G_n$, and therefore

$$\sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \geq \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n). \tag{5.209}$$

It follows by Proposition 5.45 that the quantity on the right-hand side of (5.209) is greater than $1 - \delta$ for sufficiently large values of n. Therefore, for sufficiently large values of n it holds that the coding scheme (f_n, g_n) is an (n, α, δ)-coding scheme, which proves the first statement of the theorem.

Now assume that $\alpha < \mathrm{H}(p)$, let a coding scheme of the form (5.205) be fixed for each n, and let $G_n \subseteq \Sigma^n$ be as defined in (5.208). It must hold that

$$|G_n| \leq 2^m = 2^{\lfloor \alpha n \rfloor} \tag{5.210}$$

for each n, as the coding scheme cannot be correct for two or more distinct strings that map to the same encoding. To complete the proof, it suffices to prove that

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) = 0. \tag{5.211}$$

Toward this goal, observe that for every $n \in \mathbb{N}$ and $\varepsilon > 0$ it holds that

$$G_n \subseteq (\Sigma^n \setminus T_{n,\varepsilon}) \cup (G_n \cap T_{n,\varepsilon}), \tag{5.212}$$

and therefore, by the union bound, one has

$$\begin{aligned} &\sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \\ &\leq \left(1 - \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n)\right) + 2^{-n(\mathrm{H}(p) - \varepsilon)} \, |G_n|. \end{aligned} \tag{5.213}$$

Choosing $\varepsilon > 0$ so that $\alpha < \mathrm{H}(p) - \varepsilon$, one has

$$\lim_{n \to \infty} 2^{-n(\mathrm{H}(p) - \varepsilon)} |G_n| = 0. \tag{5.214}$$

As Proposition 5.45 implies that

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n) = 1, \tag{5.215}$$

it follows that (5.211) holds, which completes the proof. □

5.3.2 Quantum source coding

There is a natural way to formulate a quantum analogue of classical source coding, which is as follows. It is assumed that a source produces a sequence of registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$, for some choice of a positive integer n, with all of these registers sharing a common classical state set Σ. Moreover, for some choice of a density operator $\rho \in \mathrm{D}(\mathbb{C}^{\Sigma})$, it is assumed that the state of the compound register $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ produced by the source is given by

$$\rho^{\otimes n} = \rho \otimes \cdots \otimes \rho \quad (n \text{ times}). \tag{5.216}$$

That is, the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are independent, and each is in a state that is described by ρ. The quantum information stored in these registers is to be encoded and decoded in a similar way to the classical setting, through the use of quantum channels rather than deterministic encoding and decoding functions.

Quantum coding schemes

A *quantum coding scheme* consists of a pair of channels (Φ, Ψ); the channel Φ represents the encoding process and Ψ represents the decoding process. The encoding channel Φ transforms $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ into $(\mathsf{Y}_1, \ldots, \mathsf{Y}_m)$, for some choice of an integer m, where $\mathsf{Y}_1, \ldots, \mathsf{Y}_m$ are registers having classical sets equal to the binary alphabet $\Gamma = \{0, 1\}$. In other words, each register Y_k represents a qubit. The decoding channel Ψ transforms $(\mathsf{Y}_1, \ldots, \mathsf{Y}_m)$ back into $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$.

The desired property of such a scheme is for the composition $\Psi\Phi$ to act trivially, or nearly trivially, on the compound register $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$, under the assumption that the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are independent and each in the state ρ as suggested above. It must be stressed that it is not sufficient to require that the state of $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ be close to $\rho^{\otimes n}$ after the decoding channel is applied—this would be a trivial requirement failing to recognize that there might initially be correlations among $\mathsf{X}_1, \ldots, \mathsf{X}_n$ and one or more other registers that must be respected by coding process. Indeed, for any complex Euclidean space $\mathcal{Z}$ and a state $\sigma \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \otimes \mathcal{Z})$ satisfying

$$\sigma[\mathsf{X}_1, \ldots, \mathsf{X}_n] = \rho^{\otimes n}, \tag{5.217}$$

it is required of a good coding scheme that the state $\big(\Psi\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\sigma)$ is approximately equal to σ.

The particular notion of approximate equality that will be considered is based on the fidelity function. This is a convenient choice, as it allows for the utilization of the closed-form expression of the channel fidelity given by Proposition 3.34. One could alternatively use the trace distance in place of the fidelity function, but this would not change the asymptotic behavior of the sorts of quantum coding schemes considered in this section, as the Fuchs–van de Graaf inequalities (Theorem 3.36) directly imply.

In accordance with the discussion above, quantum coding schemes are to be defined more precisely as follows. Let Σ be an alphabet, let $\rho \in \mathrm{D}(\mathbb{C}^{\Sigma})$ be a density operator, and let n be a positive integer. Also let $\Gamma = \{0,1\}$ denote the binary alphabet, let $\alpha > 0$ and $\delta \in (0,1)$ be real numbers, let $m = \lfloor \alpha n \rfloor$, and let $\mathcal{X}_1 = \mathbb{C}^{\Sigma}$, $\ldots$, $\mathcal{X}_n = \mathbb{C}^{\Sigma}$ and $\mathcal{Y}_1 = \mathbb{C}^{\Gamma}$, $\ldots$, $\mathcal{Y}_m = \mathbb{C}^{\Gamma}$ be complex Euclidean spaces. A pair of channels

$$\begin{aligned} \Phi &\in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) \\ \Psi &\in \mathrm{C}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n) \end{aligned} \tag{5.218}$$

is an (n, α, δ)-*quantum coding scheme* for ρ if and only if it holds that

$$\mathrm{F}\bigl(\Psi\Phi, \rho^{\otimes n}\bigr) > 1 - \delta, \tag{5.219}$$

for $\mathrm{F}\bigl(\Psi\Phi, \rho^{\otimes n}\bigr)$ denoting the channel fidelity of $\Psi\Phi$ with respect to $\rho^{\otimes n}$ (q.v. Section 3.2.3).

Schumacher's quantum source coding theorem

The following theorem is a quantum analogue to Shannon's source coding theorem (Theorem 5.43), establishing conditions on α for which quantum coding schemes exist.

Theorem 5.47 (Schumacher)**.** *Let Σ be an alphabet, let $\rho \in \mathrm{D}(\mathbb{C}^{\Sigma})$ be a density operator, and let $\alpha > 0$ and $\delta \in (0,1)$ be real numbers. The following statements hold:*

1. *If $\alpha > \mathrm{H}(\rho)$, then there exists an (n, α, δ)-quantum coding scheme for ρ for all but finitely many choices of $n \in \mathbb{N}$.*
2. *If $\alpha < \mathrm{H}(\rho)$, then there exists an (n, α, δ)-quantum coding scheme for ρ for at most finitely many choices of $n \in \mathbb{N}$.*

Proof. By the spectral theorem, one may write

$$\rho = \sum_{a \in \Sigma} p(a)\, u_a u_a^*, \tag{5.220}$$

for some choice of a probability vector $p \in \mathcal{P}(\Sigma)$ and an orthonormal basis $\{u_a : a \in \Sigma\}$ of $\mathbb{C}^\Sigma$. The association of the eigenvectors and eigenvalues of ρ with the elements of Σ may be chosen arbitrarily, and is assumed to be fixed for the remainder of the proof. By the definition of the von Neumann entropy, it holds that $\mathrm{H}(\rho) = \mathrm{H}(p)$.

Assume first that $\alpha > \mathrm{H}(\rho)$, and choose $\varepsilon > 0$ to be sufficiently small so that $\alpha > \mathrm{H}(\rho) + 2\varepsilon$. Along similar lines to the proof of Theorem 5.43, a quantum coding scheme (Φ_n, Ψ_n) of the form

$$\begin{aligned} \Phi_n &\in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) \\ \Psi_n &\in \mathrm{C}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n) \end{aligned} \tag{5.221}$$

will be defined for every $n > 1/\varepsilon$, where $m = \lfloor \alpha n \rfloor$. It will then be shown that (Φ_n, Ψ_n) is an (n, α, δ)-quantum coding scheme for sufficiently large values of n.

For a given choice of $n > 1/\varepsilon$, the quantum coding scheme (Φ_n, Ψ_n) is defined as follows. First, consider the set of ε-typical strings

$$T_{n,\varepsilon} = T_{n,\varepsilon}(p) \subseteq \Sigma^n \tag{5.222}$$

associated with the probability vector p, and define a projection operator $\Pi_{n,\varepsilon} \in \mathrm{Proj}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ as follows:

$$\Pi_{n,\varepsilon} = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} u_{a_1} u_{a_1}^* \otimes \cdots \otimes u_{a_n} u_{a_n}^*. \tag{5.223}$$

The subspace upon which this operator projects is the *ε-typical subspace* of $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ with respect to ρ. Notice that

$$\left\langle \Pi_{n,\varepsilon}, \rho^{\otimes n} \right\rangle = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} p(a_1) \cdots p(a_n). \tag{5.224}$$

Now, by Shannon's source coding theorem (or, to be more precise, the proof of that theorem given in the previous subsection), there exists a classical coding scheme (f_n, g_n) for p that satisfies

$$g_n(f_n(a_1 \cdots a_n)) = a_1 \cdots a_n \tag{5.225}$$

for every ε-typical string $a_1 \cdots a_n \in T_{n,\varepsilon}$. Define a linear operator of the form

$$A_n \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) \tag{5.226}$$

as follows:

$$A_n = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} e_{f_n(a_1 \cdots a_n)} (u_{a_1} \otimes \cdots \otimes u_{a_n})^*. \tag{5.227}$$

Finally, define channels Φ_n and Ψ_n of the form (5.221) as

$$\Phi(X) = A_n X A_n^* + \langle \mathbb{1} - A_n^* A_n, X \rangle \, \sigma \tag{5.228}$$

$$\Psi(Y) = A_n^* Y A_n + \langle \mathbb{1} - A_n A_n^*, Y \rangle \, \xi \tag{5.229}$$

for all $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ and $Y \in \mathrm{L}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m)$, for density operators $\sigma \in \mathrm{D}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m)$ and $\xi \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ chosen arbitrarily.

It remains to prove that (Φ_n, Ψ_n) is an (n, α, δ)-quantum coding scheme for sufficiently large values of n. From the expressions (5.228) and (5.229) it follows that there must exist a Kraus representation of the channel $\Psi_n \Phi_n$ having the form

$$(\Psi_n \Phi_n)(X) = (A_n^* A_n) X (A_n^* A_n)^* + \sum_{k=1}^{N} B_{n,k} X B_{n,k}^* \tag{5.230}$$

for some choice of an integer N and a collection of operators $B_{n,1}, \ldots, B_{n,N}$, which will have no effect on the analysis that follows. By Proposition 3.34, it therefore holds that

$$\mathrm{F}\big(\Psi_n \Phi_n, \rho^{\otimes n}\big) \geq \big\langle \rho^{\otimes n}, A_n^* A_n \big\rangle = \big\langle \rho^{\otimes n}, \Pi_{n,\varepsilon} \big\rangle. \tag{5.231}$$

As

$$\lim_{n \to \infty} \big\langle \Pi_{n,\varepsilon}, \rho^{\otimes n} \big\rangle = 1, \tag{5.232}$$

it follows that (Φ_n, Ψ_n) is an (n, α, δ)-quantum coding scheme for all sufficiently large n, which proves the first statement in the theorem.

Now assume that $\alpha < \mathrm{H}(\rho)$, and suppose that Φ_n and Ψ_n are arbitrary channels of the form (5.221) for each $n \in \mathbb{N}$. It will be proved that, for any choice of $\delta \in (0, 1)$, the pair (Φ_n, Ψ_n) fails to be an (n, α, δ) quantum coding scheme for all sufficiently large values of n.

Fix any choice of $n \in \mathbb{N}$, and let

$$\Phi_n(X) = \sum_{k=1}^{N} A_k X A_k^* \quad \text{and} \quad \Psi_n(X) = \sum_{k=1}^{N} B_k X B_k^* \tag{5.233}$$

be Kraus representations of Φ_n and Ψ_n, where

$$\begin{aligned} A_1, \ldots, A_N &\in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m), \\ B_1, \ldots, B_N &\in \mathrm{L}(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n). \end{aligned} \tag{5.234}$$

(The assumption that both representations have the same number of Kraus operators is made only for notational convenience. This assumption causes no loss of generality; one may include the zero operator as a Kraus operator for either channel any desired number of times.) It follows that

$$(\Psi_n \Phi_n)(X) = \sum_{1 \leq j,k \leq N} (B_k A_j) X (B_k A_j)^* \tag{5.235}$$

is a Kraus representation of the composition $\Psi_n \Phi_n$. For the purposes of this analysis, the key aspect of this Kraus representation is that

$$\operatorname{rank}(B_k A_j) \leq \dim(\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m) = 2^m \tag{5.236}$$

for all choices of $j, k \in \{1, \ldots, N\}$. Indeed, for each $k \in \{1, \ldots, N\}$, one may choose a projection operator $\Pi_k \in \operatorname{Proj}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ with $\operatorname{rank}(\Pi_k) \leq 2^m$ such that $\Pi_k B_k = B_k$. Therefore,

$$\begin{aligned} \mathrm{F}\big(\Psi_n \Phi_n, \rho^{\otimes n}\big)^2 &= \sum_{1 \leq j,k \leq N} \big|\big\langle B_k A_j, \rho^{\otimes n} \big\rangle\big|^2 \\ &= \sum_{1 \leq j,k \leq N} \big|\big\langle \Pi_k B_k A_j, \rho^{\otimes n} \big\rangle\big|^2 \\ &= \sum_{1 \leq j,k \leq N} \Big|\Big\langle B_k A_j \sqrt{\rho^{\otimes n}}, \Pi_k \sqrt{\rho^{\otimes n}} \Big\rangle\Big|^2 \\ &\leq \sum_{1 \leq j,k \leq N} \operatorname{Tr}\Big(B_k A_j \rho^{\otimes n} A_j^* B_k^*\Big) \big\langle \Pi_k, \rho^{\otimes n} \big\rangle, \end{aligned} \tag{5.237}$$

where the inequality follows from the Cauchy–Schwarz inequality. As each Π_k has rank bounded by 2^m, it follows that

$$\big\langle \Pi_k, \rho^{\otimes n} \big\rangle \leq \sum_{i=1}^{2^m} \lambda_i(\rho^{\otimes n}) = \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \tag{5.238}$$

for some subset $G_n \subseteq \Sigma^n$ having size at most 2^m. As the channel $\Psi_n \Phi_n$ is trace-preserving, it holds that

$$\sum_{1 \leq j,k \leq N} \operatorname{Tr}\Big(B_k A_j \rho^{\otimes n} A_j^* B_k^*\Big) = 1, \tag{5.239}$$

and, moreover, one has that each term in this sum is nonnegative. The final expression of (5.237) is therefore equal to a convex combination of values, each of which is bounded as in (5.238), which implies that

$$\mathrm{F}\bigl(\Psi_n \Phi_n, \rho^{\otimes n}\bigr)^2 \leq \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) \tag{5.240}$$

for some subset $G_n \subseteq \Sigma^n$ having size at most 2^m.

Finally, reasoning precisely as in the proof of Theorem 5.43, one has that the assumption $\alpha < \mathrm{H}(\rho) = \mathrm{H}(p)$ implies that

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in G_n} p(a_1) \cdots p(a_n) = 0 \tag{5.241}$$

for any choice of sets $G_n \subseteq \Sigma^n$ having size bounded by 2^m. This implies that, for any fixed choice of $\delta \in (0,1)$, the pair (Φ_n, Ψ_n) fails to be a (n, α, δ) quantum coding scheme for all but finitely many values of n. □

5.3.3 Encoding classical information into quantum states

The final type of source coding to be discussed in this section is one in which classical information is encoded into a quantum state, and then decoded by means of a measurement. The following scenario represents one abstraction of this task.

Scenario 5.48. Let X and Z be classical registers having classical state sets Σ and Γ, respectively, and let Y be a quantum register. Also let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let $\{\rho_a : a \in \Sigma\} \subset \mathrm{D}(\mathcal{Y})$ be a collection of density operators, and let $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be a measurement.

Alice obtains an element $a \in \Sigma$, stored in the register X, that has been randomly generated by a source according to the probability vector p. She prepares Y in the state ρ_a and sends Y to Bob. Bob measures Y with respect to the measurement μ, and stores the outcome of this measurement in the classical register Z. This measurement outcome represents information that Bob has obtained regarding the classical state of X.

It is natural to consider the situation in which $\Gamma = \Sigma$ in this scenario, and to imagine that Bob aims to recover the symbol stored in Alice's register X; this is essentially the state discrimination problem discussed in Section 3.1.2. In

the discussion that follows, however, it will not be taken as an assumption that this is necessarily Bob's strategy.

Assuming that Alice and Bob operate as described in Scenario 5.48, the pair (X, Z) will be left in the probabilistic state $q \in \mathcal{P}(\Sigma \times \Gamma)$ defined by

$$q(a, b) = p(a)\langle \mu(b), \rho_a \rangle \tag{5.242}$$

for every pair $(a, b) \in \Sigma \times \Gamma$. For an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined as

$$\eta(a) = p(a)\,\rho_a \tag{5.243}$$

for each $a \in \Sigma$, the probability vector q may equivalently be expressed as

$$q(a, b) = \langle \mu(b), \eta(a) \rangle \tag{5.244}$$

for each $(a, b) \in \Sigma \times \Gamma$.

One fundamental question regarding this scenario is the following: How much information can Bob's register Z contains about the state of Alice's register X? A theorem known as *Holevo's theorem* establishes an upper bound on this amount of information, as represented by the mutual information between Alice's register X and Bob's register Z. Holevo's theorem is phrased in terms of two functions of the ensemble η, the *accessible information* and the *Holevo information*, which are introduced below.

Accessible information

With Scenario 5.48 and the discussion above in mind, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble, let $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be a measurement, and let $q \in \mathcal{P}(\Sigma \times \Gamma)$ be the probability vector defined as in (5.244), representing a probabilistic state of the pair of classical registers (X, Z). The notation $\mathrm{I}_\mu(\eta)$ will denote the mutual information between X and Z, with respect to a probabilistic state defined in this way, so that

$$\mathrm{I}_\mu(\eta) = \mathrm{H}(q[\mathsf{X}]) + \mathrm{H}(q[\mathsf{Z}]) - \mathrm{H}(q) = \mathrm{D}\big(q \,\big\|\, q[\mathsf{X}] \otimes q[\mathsf{Z}]\big). \tag{5.245}$$

Now suppose that the ensemble η is fixed, while the measurement μ is unconstrained. The *accessible information* $\mathrm{I}_{\mathrm{acc}}(\eta)$ of the ensemble η is defined as the supremum value, ranging over all possible choices of a measurement μ, that may be obtained in this way. That is,

$$\mathrm{I}_{\mathrm{acc}}(\eta) = \sup_{\mu} \mathrm{I}_\mu(\eta), \tag{5.246}$$

where the supremum is over all choices of an alphabet Γ and a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$.

Although it is not necessarily apparent from its definition, the accessible information $\mathrm{I}_{\mathrm{acc}}(\eta)$ of an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ is indeed achieved by some choice of an alphabet Γ and a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$. The following lemma is useful for establishing this fact.

Lemma 5.49. *Let Σ and Γ be alphabets, let $\mathcal{Y}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble of states. Also let $\mu_0, \mu_1 : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be measurements and let $\lambda \in [0, 1]$ be a real number. It holds that*

$$\mathrm{I}_{\lambda\mu_0 + (1-\lambda)\mu_1}(\eta) \leq \lambda\, \mathrm{I}_{\mu_0}(\eta) + (1-\lambda)\, \mathrm{I}_{\mu_1}(\eta). \tag{5.247}$$

Proof. Let X and Z be classical registers having classical state sets Σ and Γ, respectively. Define a probability vector $p \in \mathcal{P}(\Sigma)$ as

$$p(a) = \mathrm{Tr}(\eta(a)) \tag{5.248}$$

for all $a \in \Sigma$. Also define probability vectors $q_0, q_1 \in \mathcal{P}(\Sigma \times \Gamma)$, representing probabilistic states of the pair (X, Z), as

$$q_0(a, b) = \langle \mu_0(b), \eta(a) \rangle \quad \text{and} \quad q_1(a, b) = \langle \mu_1(b), \eta(a) \rangle \tag{5.249}$$

for all $(a, b) \in \Sigma \times \Gamma$. By the joint convexity of the relative entropy function, it holds that

$$\begin{aligned}
&\mathrm{I}_{\lambda\mu_0 + (1-\lambda)\mu_1}(\eta) \\
&\quad = \mathrm{D}\big(\lambda q_0 + (1-\lambda) q_1 \,\big\|\, p \otimes (\lambda q_0[\mathsf{Z}] + (1-\lambda) q_1[\mathsf{Z}])\big) \\
&\quad \leq \lambda\, \mathrm{D}\big(q_0 \,\big\|\, p \otimes q_0[\mathsf{Z}]\big) + (1-\lambda)\, \mathrm{D}\big(q_1 \,\big\|\, p \otimes q_1[\mathsf{Z}]\big) \\
&\quad = \lambda\, \mathrm{I}_{\mu_0}(\eta) + (1-\lambda)\, \mathrm{I}_{\mu_1}(\eta)\,,
\end{aligned} \tag{5.250}$$

as required. □

Theorem 5.50. *Let Σ be an alphabet, let $\mathcal{Y}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble of states. There exists an alphabet Γ with $|\Gamma| \leq \dim(\mathcal{Y})^2$ and a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ such that $\mathrm{I}_{\mu}(\eta) = \mathrm{I}_{\mathrm{acc}}(\eta)$.*

Proof. Let $\nu : \Delta \to \mathrm{Pos}(\mathcal{Y})$ be a measurement, for an arbitrary choice of an alphabet Δ. By Lemma 5.49, the function $\mu \mapsto \mathrm{I}_{\mu}(\eta)$ is convex on the set of all measurements of the form $\mu : \Delta \to \mathrm{Pos}(\mathcal{Y})$. As every measurement of

this form can be written as a convex combination of extremal measurements of the same form, one has that there must exist an extremal measurement $\mu : \Delta \to \mathrm{Pos}(\mathcal{Y})$ satisfying $\mathrm{I}_\mu(\eta) \geq \mathrm{I}_\nu(\eta)$. By Corollary 2.48, the assumption that $\mu : \Delta \to \mathrm{Pos}(\mathcal{Y})$ is extremal implies that

$$\big|\big\{a \in \Delta \,:\, \mu(a) \neq 0\big\}\big| \leq \dim(\mathcal{Y})^2. \tag{5.251}$$

The value $\mathrm{I}_\mu(\eta)$ does not change if μ is restricted to the alphabet

$$\Gamma = \big\{a \in \Delta \,:\, \mu(a) \neq 0\big\}, \tag{5.252}$$

and therefore one has that there must exist a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$, for Γ satisfying $|\Gamma| \leq \dim(\mathcal{Y})^2$, such that $\mathrm{I}_\mu(\eta) \geq \mathrm{I}_\nu(\eta)$.

It follows that $\mathrm{I}_{\mathrm{acc}}(\eta)$ is equal to the supremum value of $\mathrm{I}_\mu(\eta)$, ranging over all measurements μ having at most $\dim(\mathcal{Y})^2$ measurement outcomes. The quantity $\mathrm{I}_\mu(\eta)$ is invariant under renaming the measurement outcomes of μ, so there is no loss of generality in restricting this supremum to the set of measurements having a single set Γ of measurement outcomes satisfying $|\Gamma| = \dim(\mathcal{Y})^2$. The supremum is therefore taken over a compact set, from which it follows that there exists a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ for which the supremum value is achieved, which completes the proof. □

The Holevo information

Again with Scenario 5.48 in mind, let X be a classical register, let Σ be the classical state set of X, let Y be a quantum register, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble. As described in Section 2.2.3, one associates the classical-quantum state

$$\sigma = \sum_{a \in \Sigma} E_{a,a} \otimes \eta(a) \tag{5.253}$$

of the pair (X, Y) with the ensemble η. The *Holevo information* (also called the *Holevo χ-quantity*) of the ensemble η, which is denoted $\chi(\eta)$, is defined as the quantum mutual information $\mathrm{I}(\mathsf{X} : \mathsf{Y})$ between the registers X and Y with respect to the state σ.

Under the assumption that the ensemble η is written as

$$\eta(a) = p(a)\,\rho_a \tag{5.254}$$

for each $a \in \Sigma$, for a probability vector $p \in \mathcal{P}(\Sigma)$ and a collection

$$\{\rho_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y}) \tag{5.255}$$

of states, the Holevo information of η may be calculated as follows:

$$\begin{aligned}\chi(\eta) &= \mathrm{I}(\mathsf{X}:\mathsf{Y}) \\ &= \mathrm{H}(\mathsf{X}) + \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{X},\mathsf{Y}) \\ &= \mathrm{H}(p) + \mathrm{H}\biggl(\sum_{a\in\Sigma} p(a)\,\rho_a\biggr) - \mathrm{H}\biggl(\sum_{a\in\Sigma} p(a)\,E_{a,a}\otimes\rho_a\biggr) \\ &= \mathrm{H}\biggl(\sum_{a\in\Sigma} p(a)\,\rho_a\biggr) - \sum_{a\in\Sigma} p(a)\;\mathrm{H}\bigl(\rho_a\bigr),\end{aligned} \tag{5.256}$$

where the last equality has made use of the identity (5.95). Alternatively, one may write

$$\chi(\eta) = \mathrm{H}\biggl(\sum_{a\in\Sigma}\eta(a)\biggr) - \sum_{\substack{a\in\Sigma\\ \eta(a)\neq 0}} \mathrm{Tr}(\eta(a))\;\mathrm{H}\biggl(\frac{\eta(a)}{\mathrm{Tr}(\eta(a))}\biggr), \tag{5.257}$$

or, equivalently,

$$\chi(\eta) = \mathrm{H}\biggl(\sum_{a\in\Sigma}\eta(a)\biggr) - \sum_{a\in\Sigma}\mathrm{H}(\eta(a)) + \mathrm{H}(p). \tag{5.258}$$

It follows from the concavity of the von Neumann entropy (Theorem 5.25), or by the subadditivity of von Neumann entropy (Proposition 5.10), that the Holevo information $\chi(\eta)$ is nonnegative for every ensemble η.

At an intuitive level, the Holevo information may be interpreted in the following way. When the pair of registers (X,Y) is in the classical-quantum state σ as described above, and the register Y is considered in isolation, its von Neumann entropy is given by

$$\mathrm{H}(\mathsf{Y}) = \mathrm{H}\biggl(\sum_{a\in\Sigma} p(a)\rho_a\biggr). \tag{5.259}$$

If one learns the classical state $a \in \Sigma$ of X, then from their perspective the von Neumann entropy of Y drops to $\mathrm{H}(\rho_a)$. The Holevo information $\chi(\eta)$ may therefore be viewed as representing the average decrease in the von Neumann entropy of Y that is expected when one learns the classical state of X.

It cannot be said that the Holevo information is convex in general, but the following proposition provides two conditions under which it is. The proof follows a similar argument to the proof of Lemma 5.49.

Proposition 5.51. *Let $\mathcal{Y}$ be a complex Euclidean space, let Σ be an alphabet, and let $\eta_0 : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ and $\eta_1 : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be ensembles of states. Suppose further that at least one of the following two conditions is satisfied:*

1. *The ensembles η_0 and η_1 have the same average state:*
$$\sum_{a\in\Sigma} \eta_0(a) = \rho = \sum_{a\in\Sigma} \eta_1(a), \tag{5.260}$$
for some choice of $\rho \in \mathrm{D}(\mathcal{Y})$.
2. *The ensembles η_0 and η_1 yield equal probability distributions, over possibly different states:*
$$\mathrm{Tr}(\eta_0(a)) = p(a) = \mathrm{Tr}(\eta_1(a)) \tag{5.261}$$
for each $a \in \Sigma$, for some choice of a probability vector $p \in \mathcal{P}(\Sigma)$.

For every real number $\lambda \in [0,1]$, it holds that
$$\chi(\lambda\eta_0 + (1-\lambda)\eta_1) \leq \lambda\chi(\eta_0) + (1-\lambda)\chi(\eta_1). \tag{5.262}$$

Proof. Let $\mathcal{X} = \mathbb{C}^\Sigma$, let X and Y be registers corresponding to the spaces $\mathcal{X}$ and $\mathcal{Y}$, and define classical-quantum states $\sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X}\otimes\mathcal{Y})$ as
$$\sigma_0 = \sum_{a\in\Sigma} E_{a,a} \otimes \eta_0(a) \quad \text{and} \quad \sigma_1 = \sum_{a\in\Sigma} E_{a,a} \otimes \eta_1(a). \tag{5.263}$$
For a given choice of $\lambda \in [0,1]$, define $\sigma = \lambda\sigma_0 + (1-\lambda)\sigma_1$. The Holevo information of the ensembles η_0, η_1, and $\lambda\eta_0 + (1-\lambda)\eta_1$ may be expressed as follows:
$$\begin{aligned} \chi(\eta_0) &= \mathrm{H}\big(\sigma_0\big\|\sigma_0[\mathsf{X}]\otimes\sigma_0[\mathsf{Y}]\big), \\ \chi(\eta_1) &= \mathrm{H}\big(\sigma_1\big\|\sigma_1[\mathsf{X}]\otimes\sigma_1[\mathsf{Y}]\big), \end{aligned} \tag{5.264}$$
and
$$\chi(\lambda\eta_0 + (1-\lambda)\eta_1) = \mathrm{H}\big(\sigma\big\|\sigma[\mathsf{X}]\otimes\sigma[\mathsf{Y}]\big). \tag{5.265}$$

Under the first condition in the statement of the proposition, it holds that $\sigma_0[\mathsf{Y}] = \sigma_1[\mathsf{Y}] = \sigma[\mathsf{Y}] = \rho$. In this case, the inequality (5.262) is equivalent to
$$\mathrm{H}\big(\sigma\big\|\sigma[\mathsf{X}]\otimes\rho\big) \leq \lambda\,\mathrm{H}\big(\sigma_0\big\|\sigma_0[\mathsf{X}]\otimes\rho\big) + (1-\lambda)\,\mathrm{H}\big(\sigma_1\big\|\sigma_1[\mathsf{X}]\otimes\rho\big), \tag{5.266}$$
which follows from the joint convexity of the quantum relative entropy function (Corollary 5.36). Under the second condition in the statement of the proposition, one has $\sigma_0[\mathsf{X}] = \sigma_1[\mathsf{X}] = \sigma[\mathsf{X}] = \mathrm{Diag}(p)$. Exchanging the roles of X and Y from the first condition, one has that the the proof follows by similar reasoning. □

Holevo's theorem

The next theorem, known as *Holevo's theorem*, establishes that the accessible information is upper-bounded by the Holevo information, for all ensembles of states.

Theorem 5.52 (Holevo's theorem)**.** *Let Σ be an alphabet, let $\mathcal{Y}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble of states. It holds that $I_{\mathrm{acc}}(\eta) \leq \chi(\eta)$.*

Proof. Let X be a classical register having classical state set Σ and let Y be a register whose associated complex Euclidean space is $\mathcal{Y}$. Define a state $\sigma \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ as

$$\sigma = \sum_{a \in \Sigma} E_{a,a} \otimes \eta(a), \tag{5.267}$$

and suppose that the pair (X, Y) is in the state σ. It holds that

$$\chi(\eta) = \mathrm{D}\big(\sigma \big\| \sigma[\mathsf{X}] \otimes \sigma[\mathsf{Y}]\big). \tag{5.268}$$

Next, let Γ be an alphabet, let Z be a classical register having classical state set Γ, and let $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be a measurement. Define a channel $\Phi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ as

$$\Phi(Y) = \sum_{b \in \Gamma} \langle \mu(b), Y \rangle E_{b,b} \tag{5.269}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$, which is the quantum-to-classical channel associated with the measurement μ, and consider the situation in which Y is transformed into Z by means of Φ. One has that

$$(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\sigma) = \sum_{a \in \Sigma} \sum_{b \in \Gamma} \big\langle \mu(b), \eta(a) \big\rangle E_{a,a} \otimes E_{b,b} = \mathrm{Diag}(q), \tag{5.270}$$

for $q \in \mathcal{P}(\Sigma \times \Gamma)$ being the probability vector defined as

$$q(a,b) = \big\langle \mu(b), \eta(a) \big\rangle \tag{5.271}$$

for all $a \in \Sigma$ and $b \in \Gamma$. It follows that

$$\begin{aligned} \mathrm{I}_{\mu}(\eta) &= \mathrm{D}(q \| q[\mathsf{X}] \otimes q[\mathsf{Z}]) \\ &= \mathrm{D}\big((\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\sigma) \big\| (\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Phi)(\sigma[\mathsf{X}] \otimes \sigma[\mathsf{Y}])\big), \end{aligned} \tag{5.272}$$

and therefore $\mathrm{I}_{\mu}(\eta) \leq \chi(\eta)$, as the quantum relative entropy does not increase under the action of a channel (by Theorem 5.38). As this bound holds for all measurements μ, the theorem follows. □

For every collection of density operators $\{\rho_a : a \in \Sigma\} \subset \mathrm{D}(\mathcal{Y})$ and every probability vector $p \in \mathcal{P}(\Sigma)$, it holds that

$$\begin{aligned} \mathrm{H}\left(\sum_{a\in\Sigma} p(a)\rho_a\right) - \sum_{a\in\Sigma} p(a)\,\mathrm{H}(\rho_a) \\ \leq \mathrm{H}\left(\sum_{a\in\Sigma} p(a)\rho_a\right) \leq \log(\dim(\mathcal{Y})), \end{aligned} \tag{5.273}$$

and therefore the Holevo information of every ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ is upper-bounded by $\log(\dim(\mathcal{Y}))$. The following corollary of Theorem 5.52 is a consequence of this observation.

Corollary 5.53. *Let Σ be an alphabet, let $\mathcal{Y}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ be an ensemble of states. It holds that $\mathrm{I}_{\mathrm{acc}}(\eta) \leq \log(\dim(\mathcal{Y}))$.*

Although this is indeed a simple corollary to Theorem 5.52, it nevertheless establishes the following conceptually important fact: if two individuals share no prior correlations or shared resources, and one individual sends the other a quantum register of a given dimension n, then no more than $\log(n)$ bits of classical information will have been transmitted through this process.

Quantum random access codes

An interesting variation of source coding involves the notion of a *quantum random access code*. This is a coding scheme in which a sequence of classical symbols is encoded into a quantum state in such a way that one may obtain information about just one of the encoded symbols, chosen arbitrarily by the individual performing the decoding operation. The following scenario provides an abstraction of this type of scheme.

Scenario 5.54. Let Σ and Γ be alphabets, let n be a positive integer, let $\mathsf{X}_1, \ldots, \mathsf{X}_n$ be classical registers, each having classical state set Σ, let Z be a classical register having classical state set Γ, and let Y be a quantum register. Also let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let

$$\left\{\rho_{a_1\cdots a_n} : a_1 \cdots a_n \in \Sigma^n\right\} \subset \mathrm{D}(\mathcal{Y}) \tag{5.274}$$

be a collection of states indexed by Σ^n, and let $\mu_1, \ldots, \mu_n : \Gamma \to \mathrm{Pos}(\mathcal{Y})$ be measurements.

Alice obtains the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$, which have been independently prepared by a source, with p being the probabilistic state of each of these registers. She observes the classical state $a_1 \cdots a_n \in \Sigma^n$ of $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$, and prepares the register Y in the state $\rho_{a_1 \cdots a_n}$, which is then sent to Bob. Bob selects an index $k \in \{1, \ldots, n\}$ of his choice, measures Y with respect to the measurement μ_k, and stores the outcome in the classical register Z. The classical state of Z represents the information Bob has obtained regarding the classical state of X_k.

The following example describes an instance of this scenario in which Alice encodes two classical bits into one qubit in such a way that Bob can recover the encoded bit of his choice with a reasonably high probability of success.

Example 5.55. Let $\Sigma = \{0, 1\}$ denote the binary alphabet. For every real number θ, define a density operator $\sigma(\theta) \in \mathrm{D}(\mathbb{C}^\Sigma)$ as

$$\sigma(\theta) = \begin{pmatrix} \cos^2(\theta) & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \sin^2(\theta) \end{pmatrix}, \tag{5.275}$$

and observe that each of these operators is a rank one projection.

Alice obtains two classical registers X_1 and X_2, both having classical state set Σ. It is to be assumed that the probabilistic states of these registers are independent and uniformly distributed. She encodes the classical state $(a_1, a_2) \in \Sigma \times \Sigma$ of the pair $(\mathsf{X}_1, \mathsf{X}_2)$ into the quantum state $\rho_{a_1 a_2} \in \mathrm{D}(\mathbb{C}^\Sigma)$ defined as

$$\begin{aligned} \rho_{00} &= \sigma(\pi/8), \quad & \rho_{10} &= \sigma(3\pi/8), \\ \rho_{01} &= \sigma(7\pi/8), \quad & \rho_{11} &= \sigma(5\pi/8). \end{aligned} \tag{5.276}$$

Bob receives the qubit $\rho_{a_1 a_2}$ from Alice, and decides whether he wishes to learn the classical state a_1 of X_1 or the classical state a_2 of X_2. If he wishes to learn a_1, he measures the qubit with respect to the measurement μ_1 defined as

$$\mu_1(0) = \sigma(0) \quad \text{and} \quad \mu_1(1) = \sigma(\pi/2). \tag{5.277}$$

If he wishes to learn a_2, he measures the qubit with respect to the measurement μ_2 defined as

$$\mu_2(0) = \sigma(\pi/4) \quad \text{and} \quad \mu_2(1) = \sigma(3\pi/4). \tag{5.278}$$

Using the formula

$$\langle \sigma(\phi), \sigma(\theta) \rangle = \cos^2(\phi - \theta), \tag{5.279}$$

one concludes from a case analysis that, if Bob measures $\rho_{a_1 a_2}$ with respect to the measurement μ_k, he will obtain the measurement outcome a_k with probability

$$\cos^2(\pi/8) \approx 0.85 \tag{5.280}$$

in all cases.

With Scenario 5.54 in mind, one may define a *quantum random access code* for a given choice of a positive integer n and a probability vector $p \in \mathcal{P}(\Sigma)$ as consisting of two objects: the first is the collection of density operators

$$\{\rho_{a_1 \cdots a_n} : a_1 \cdots a_n \in \Sigma^n\} \subseteq \mathrm{D}(\mathcal{Y}) \tag{5.281}$$

representing the encodings of the possible sequences $a_1 \cdots a_n \in \Sigma^n$, and the second is the sequence of measurements

$$\mu_1, \ldots, \mu_n : \Gamma \to \mathrm{Pos}(\mathcal{Y}) \tag{5.282}$$

that reveal information concerning one of the initial registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$.

The amount of information revealed by such a quantum random access code may be represented by a vector $(\alpha_1, \ldots, \alpha_n)$, where α_k represents the mutual information between X_k and Z, conditioned on the measurement μ_k having been performed and the outcome of that measurement stored in Z. The vector $(\alpha_1, \ldots, \alpha_n)$ may be defined in more precise terms as follows. First, one defines an ensemble $\eta : \Sigma^n \to \mathrm{Pos}(\mathcal{Y})$ as

$$\eta(a_1 \cdots a_n) = p(a_1) \cdots p(a_n)\, \rho_{a_1 \cdots a_n} \tag{5.283}$$

for each $a_1 \cdots a_n \in \Sigma^n$. Then, for each $k \in \{1, \ldots, n\}$, one defines

$$\alpha_k = \mathrm{I}(\mathsf{X}_k : \mathsf{Z}), \tag{5.284}$$

where the mutual information is defined with respect to the probabilistic state $q_k \in \mathcal{P}(\Sigma^n \times \Gamma)$ of the compound register $(\mathsf{X}_1, \ldots, \mathsf{X}_n, \mathsf{Z})$ given by

$$q_k(a_1 \cdots a_n, b) = \langle \mu_k(b), \eta(a_1 \cdots a_n) \rangle \tag{5.285}$$

for each $a_1 \cdots a_n \in \Sigma^n$ and $b \in \Gamma$.

Nayak's theorem

Although Example 5.55 suggests a potential for quantum random access codes to provide significant advantages over classical coding schemes, it is a false impression. The following theorem demonstrates that quantum random access codes are strongly limited in their capabilities.

Theorem 5.56 (Nayak's theorem)**.** *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let n be a positive integer. Also let $\mathcal{Y}$ be a complex Euclidean space, let Γ be an alphabet, and let*

$$\{\rho_{a_1\cdots a_n} : a_1\cdots a_n \in \Sigma^n\} \subset \mathrm{D}(\mathcal{Y}) \quad \text{and} \quad \mu_1,\ldots,\mu_n : \Gamma \to \mathrm{Pos}(\mathcal{Y}) \tag{5.286}$$

be a quantum random access code for p. If $(\alpha_1,\ldots,\alpha_n)$ is a vector representing the amount of information revealed by this code for the distribution p, in the manner defined above, then it must hold that

$$\sum_{k=1}^{n} \alpha_k \leq \chi(\eta) \tag{5.287}$$

for $\eta : \Sigma^n \to \mathrm{Pos}(\mathcal{Y})$ being the ensemble defined by

$$\eta(a_1\cdots a_n) = p(a_1)\cdots p(a_n)\rho_{a_1\cdots a_n} \tag{5.288}$$

for each $a_1\cdots a_n \in \Sigma^n$.

Proof. Let $\mathsf{X}_1,\ldots,\mathsf{X}_n$ be classical registers having classical state set Σ, and let Y be a register whose associated complex Euclidean space is $\mathcal{Y}$ (as in Scenario 5.54). Let

$$\sigma = \sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n) E_{a_1,a_1}\otimes\cdots\otimes E_{a_n,a_n}\otimes\rho_{a_1\cdots a_n} \tag{5.289}$$

be the classical-quantum state of the compound register $(\mathsf{X}_1,\ldots,\mathsf{X}_n,\mathsf{Y})$ corresponding to the ensemble η. With respect to the state σ, one has that

$$\mathrm{I}(\mathsf{X}_1,\ldots,\mathsf{X}_n : \mathsf{Y}) = \chi(\eta). \tag{5.290}$$

Now, it holds that

$$\begin{aligned} &\mathrm{I}(\mathsf{X}_1,\ldots,\mathsf{X}_n : \mathsf{Y}) \\ &\quad = \mathrm{I}(\mathsf{X}_n : \mathsf{Y}) + \mathrm{I}(\mathsf{X}_1,\ldots,\mathsf{X}_{n-1} : \mathsf{X}_n,\mathsf{Y}) - \mathrm{I}(\mathsf{X}_1,\ldots,\mathsf{X}_{n-1} : \mathsf{X}_n). \end{aligned} \tag{5.291}$$

This identity (which is equivalent to an identity commonly known as the *chain rule* for quantum mutual information) holds independent of the state of these registers, and may be verified by expanding the definition of the quantum mutual information. In the particular case of the state σ, one has that

$$\mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_{n-1} : \mathsf{X}_n) = 0, \tag{5.292}$$

as the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ are independent with respect to this state. Thus,

$$\begin{aligned} \mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_n : \mathsf{Y}) &= \mathrm{I}(\mathsf{X}_n : \mathsf{Y}) + \mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_{n-1} : \mathsf{X}_n, \mathsf{Y}) \\ &\geq \mathrm{I}(\mathsf{X}_n : \mathsf{Y}) + \mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_{n-1} : \mathsf{Y}), \end{aligned} \tag{5.293}$$

where the inequality holds by Corollary 5.53. By applying this inequality recursively, one finds that

$$\mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_n : \mathsf{Y}) \geq \sum_{k=1}^{n} \mathrm{I}(\mathsf{X}_k : \mathsf{Y}). \tag{5.294}$$

Finally, one may observe that $\alpha_k \leq \mathrm{I}(\mathsf{X}_k : \mathsf{Y})$ for each $k \in \{1, \ldots, n\}$, as a consequence of Holevo's theorem (Theorem 5.52). Thus,

$$\sum_{k=1}^{n} \alpha_k \leq \mathrm{I}(\mathsf{X}_1, \ldots, \mathsf{X}_n : \mathsf{Y}) = \chi(\eta), \tag{5.295}$$

as required. □

One interesting type of quantum random access code, which includes the code suggested by Example 5.55, is one in which Σ and Γ are equal to the binary alphabet, and one aims for the classical state of the register Z to agree with X_k for whichever index $k \in \{1, \ldots, n\}$ was measured. Theorem 5.56 implies a strong limitation on schemes of this sort. The following lemma, which is a special case of an inequality known as *Fano's inequality*, is useful for analyzing this special case.

Lemma 5.57. *Let* X *and* Y *be classical registers sharing the same classical state set* $\Sigma = \{0, 1\}$*, and assume the pair* (X, Y) *is in a probabilistic state* $q \in \mathcal{P}(\Sigma \times \Sigma)$ *for which* $q[\mathsf{X}](0) = q[\mathsf{X}](1) = 1/2$ *and*

$$q(0,0) + q(1,1) = \lambda \tag{5.296}$$

for $\lambda \in [0, 1]$*. (In words, the state of* X *is uniformly distributed and* Y *and* X *agree with probability* λ*.) It holds that* $\mathrm{I}(\mathsf{X} : \mathsf{Y}) \geq 1 - \mathrm{H}(\lambda, 1 - \lambda)$*.*

Proof. Define Z to be a classical register having classical state set Σ, and let $p \in \mathcal{P}(\Sigma \times \Sigma \times \Sigma)$ be the probability vector defined as

$$p(a,b,c) = \begin{cases} q(a,b) & \text{if } c = a \oplus b \\ 0 & \text{otherwise,} \end{cases} \tag{5.297}$$

where $a \oplus b$ denotes the exclusive-OR of the binary values a and b. In words, p describes the probabilistic state of $(\mathsf{X},\mathsf{Y},\mathsf{Z})$ for which (X,Y) is distributed according to q and Z is set to the exclusive-OR of X and Y. With respect to this state, one has

$$\mathrm{H}(\mathsf{Z}) = \mathrm{H}(\lambda, 1-\lambda). \tag{5.298}$$

Moreover, it holds that

$$\mathrm{H}(\mathsf{X}|\mathsf{Y}) = \mathrm{H}(\mathsf{Z}|\mathsf{Y}), \tag{5.299}$$

as the classical states of X and Z uniquely determine one another for each fixed classical state of Y. Finally, by the subadditivity of Shannon entropy (Proposition 5.10), one has that

$$\mathrm{H}(\mathsf{Z}|\mathsf{Y}) \leq \mathrm{H}(\mathsf{Z}). \tag{5.300}$$

Consequently,

$$\begin{aligned} \mathrm{I}(\mathsf{X}:\mathsf{Y}) &= \mathrm{H}(\mathsf{X}) - \mathrm{H}(\mathsf{X}|\mathsf{Y}) = 1 - \mathrm{H}(\mathsf{Z}|\mathsf{Y}) \\ &\geq 1 - \mathrm{H}(\mathsf{Z}) = 1 - \mathrm{H}(\lambda, 1-\lambda), \end{aligned} \tag{5.301}$$

as required. □

Corollary 5.58. *Let $\Sigma = \{0,1\}$ denote the binary alphabet, let n be a positive integer, let $\mathcal{Y}$ be a complex Euclidean space, and let $\lambda \in [1/2, 1]$ be a real number. Also let*

$$\{\rho_{a_1 \cdots a_n} \,:\, a_1 \cdots a_n \in \Sigma^n\} \subset \mathrm{D}(\mathcal{Y}) \tag{5.302}$$

be a collection of density operators, and let

$$\mu_1, \ldots, \mu_n : \Sigma \to \mathrm{Pos}(\mathcal{Y}) \tag{5.303}$$

be measurements. If it holds that

$$\langle \mu_k(a_k), \rho_{a_1 \cdots a_n} \rangle \geq \lambda \tag{5.304}$$

for every choice of $a_1 \ldots a_n \in \Sigma^n$, then

$$\log(\dim(\mathcal{Y})) \geq (1 - \mathrm{H}(\lambda, 1-\lambda))n. \tag{5.305}$$

Proof. Let $p \in \mathcal{P}(\Sigma)$ be the uniform distribution and define an ensemble $\eta : \Sigma^n \to \mathrm{Pos}(\mathcal{Y})$ as

$$\eta(a_1 \cdots a_n) = p(a_1) \cdots p(a_n) \rho_{a_1 \cdots a_n} = \frac{1}{2^n} \rho_{a_1 \cdots a_n} \tag{5.306}$$

for each string $a_1 \cdots a_n \in \Sigma^n$. Let $(\alpha_1, \ldots, \alpha_n)$ be the vector representing the amount of information revealed by the quantum random access code defined by the collection $\{\rho_{a_1 \cdots a_n} : a_1 \cdots a_n \in \Sigma^n\}$ and the measurements $\mu_1, \ldots, \mu_n$ for the distribution p. By combining Lemma 5.57 with the fact that $\mathrm{H}(\alpha, 1-\alpha)$ is a decreasing function of α on the interval $[1/2, 1]$, one finds that

$$\alpha_k \geq 1 - \mathrm{H}(\lambda, 1-\lambda) \tag{5.307}$$

for every $k \in \{1, \ldots, n\}$. Therefore, by Theorem 5.56, it holds that

$$\chi(\eta) \geq (1 - \mathrm{H}(\lambda, 1-\lambda))n. \tag{5.308}$$

As the Holevo information of η is upper-bounded by $\log(\dim(\mathcal{Y}))$, the proof is complete. □

Thus, for the special type of random access code under consideration, the number of qubits required to encode a binary string of length n is linear in n, with the constant of proportionality tending to 1 as the error tolerance decreases.

5.4 Exercises

5.1. Let X, Y and Z be registers. Prove that the following inequalities hold for all states $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ of these registers.

(a) $\mathrm{I}(\mathsf{X}, \mathsf{Y} : \mathsf{Z}) + \mathrm{I}(\mathsf{Y} : \mathsf{Z}) \geq \mathrm{I}(\mathsf{X} : \mathsf{Z})$.

(b) $\mathrm{H}(\mathsf{X}, \mathsf{Y}|\mathsf{Z}) + \mathrm{H}(\mathsf{Y}|\mathsf{Z}) \geq \mathrm{H}(\mathsf{X}|\mathsf{Z}) - 2\,\mathrm{H}(\mathsf{Z})$

5.2. Let X, Y, and Z be registers.

(a) Prove that, for every state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{Z})$ of these registers, it holds that

$$\mathrm{I}(\mathsf{X}, \mathsf{Y} : \mathsf{Z}) \leq \mathrm{I}(\mathsf{Y} : \mathsf{X}, \mathsf{Z}) + 2\,\mathrm{H}(\mathsf{X}). \tag{5.309}$$

(b) Let Σ be the classical state set of X, let

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z}) \tag{5.310}$$

be a collection of density operators, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let

$$\rho = \sum_{a \in \Sigma} p(a)\, E_{a,a} \otimes \sigma_a \tag{5.311}$$

be a state of $(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$. Prove that, with respect to the state ρ, one has

$$\mathrm{I}(\mathsf{X}, \mathsf{Y} : \mathsf{Z}) \leq \mathrm{I}(\mathsf{Y} : \mathsf{X}, \mathsf{Z}) + \mathrm{H}(\mathsf{X}). \tag{5.312}$$

5.3. Let Σ be an alphabet, let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$ be a density operator, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and let

$$\{\Phi_a \,:\, a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{X}, \mathcal{Y}) \tag{5.313}$$

be a collection of channels. Define an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Z})$ as

$$\eta(a) = p(a)\big(\Phi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho) \tag{5.314}$$

for each $a \in \Sigma$. Prove that

$$\chi(\eta) \leq \mathrm{H}\left(\sum_{a \in \Sigma} p(a)\Phi_a\big(\mathrm{Tr}_{\mathcal{Z}}(\rho)\big)\right) + \sum_{a \in \Sigma} p(a)\,\mathrm{H}\big(\Phi_a\big(\mathrm{Tr}_{\mathcal{Z}}(\rho)\big)\big). \tag{5.315}$$

5.4. Let Σ be an alphabet and let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces. Also let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble, and define an ensemble $\Phi(\eta) : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ as

$$(\Phi(\eta))(a) = \Phi(\eta(a)) \tag{5.316}$$

for each $a \in \Sigma$. Prove that

$$\chi(\Phi(\eta)) \leq \chi(\eta). \tag{5.317}$$

5.5. Let X and Y be registers and let $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be states of these registers. Prove that, for every choice of $\lambda \in [0, 1]$, it holds that

$$\begin{aligned} &\mathrm{H}\big(\lambda\rho_0 + (1-\lambda)\rho_1\big) - \mathrm{H}\big(\lambda\rho_0[\mathsf{Y}] + (1-\lambda)\rho_1[\mathsf{Y}]\big) \\ &\quad \geq \lambda\big(\mathrm{H}(\rho_0) - \mathrm{H}(\rho_0[\mathsf{Y}])\big) + (1-\lambda)\big(\mathrm{H}(\rho_1) - \mathrm{H}(\rho_1[\mathsf{Y}])\big). \end{aligned} \tag{5.318}$$

(Equivalently, prove that the conditional von Neumann entropy of X given Y is a concave function of the state of these registers.)

5.6. Let X and Y be registers and let $\rho \in \mathcal{D}(\mathcal{X} \otimes \mathcal{Y})$ be a state of these registers for which it holds that

$$\rho = \sum_{a \in \Sigma} p(a) \sigma_a \otimes \xi_a,$$

for some choice of an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and two collections of states $\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X})$ and $\{\xi_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y})$.

(a) Prove that, with respect to the state ρ, it holds that $\mathrm{I}(\mathsf{X} : \mathsf{Y}) \leq \mathrm{H}(p)$.

(b) Prove that

$$\mathrm{H}(\rho) \geq \sum_{a \in \Sigma} p(a) \, \mathrm{H}(\sigma_a) + \mathrm{H}\left(\sum_{a \in \Sigma} p(a) \xi_a \right). \tag{5.319}$$

5.5 Bibliographic remarks

The Shannon entropy function was defined in Shannon's 1948 paper [191], which is generally viewed as representing the birth of information theory. Several fundamental facts were proved in that paper, including Shannon's source coding theorem (of which Theorem 5.43 is a variant) and Shannon's channel coding theorem. Shannon also defined the conditional entropy in the same paper, considered the mutual information (although not under that name), and proved that the entropy function now bearing his name is the unique function of a probability vector, up to a normalization, satisfying a few simple axioms that a measure of information and uncertainty should naturally possess. Shannon observed the similarity in form of his entropy function to the notion of entropy in statistical mechanics in his 1948 paper, and was later quoted as saying that he used the name "entropy" on the advice of von Neumann [203]. More substantive connections between these different notions of entropy have been considered by several researchers. (See, for instance, Jaynes [176].)

The relative entropy function was defined by Kullback and Leibler in 1951 [139]. Theorem 5.16 is due to Audenaert [17]. A variant of Pinsker's inequality (Theorem 5.17, but with a smaller constant factor) was proved by Pinsker [173] and later refined by others, including Csiszár and Kullback. Further information on classical information theory can be found in books on the subject, including the books of Ash [15] and Cover and Thomas [53], among many others.

The von Neumann entropy was first defined by von Neumann in a 1927 paper [213] and then investigated in greater detail in his 1932 book [217], in both cases within the context of quantum statistical mechanics. Despite Shannon's reported discussion with von Neumann regarding the Shannon entropy, there is no evidence known to suggest that von Neumann ever considered the information-theoretic aspects of the von Neumann entropy function.

The quantum relative entropy was defined by Umegaki [211] in 1962. A fact from which Klein's inequality (as stated in Proposition 5.24) may be derived was proved many years earlier by Klein [131]. Theorem 5.27 was proved by Araki and Lieb [12], who also introduced the purification method through which it is proved in the same paper. A weaker version of the Fannes–Audenaert inequality (Theorem 5.28) was proved by Fannes [71], and was later strengthened by Audenaert [17] (through a reduction to the classical result stated in Theorem 5.16, which was proved in the same paper).

Lieb's concavity theorem was proved by Lieb [146] in 1973. The precise formulation of this theorem represented by Theorem 5.32 is due to Ando [10]. Multiple proofs of this theorem are known; the proof presented in this book is an adaptation of one appearing in the book of Simon [193] with simplifications inspired by Ando's methodology in [10]. Simon attributes the central idea of his proof to Uhlmann [210]. The strong subadditivity of von Neumann entropy was first conjectured by Lanford and Robinson [143] and proved by Lieb and Ruskai [147] using Lieb's concavity theorem. The joint convexity of quantum relative entropy was proved by Lindblad [148], also through the use of Lieb's concavity theorem. The quantum Pinsker inequality (Theorem 5.41) appears in a paper of Hiai, Ohya, and Tsukada [102], and may be obtained as a special case of a more general theorem due to Uhlmann [210].

Theorem 5.47 was proved by Schumacher [185] in 1995. Holevo [105] proved his eponymous theorem (Theorem 5.52) in 1973, through a different proof than the one presented in this chapter—Holevo's proof did not make use of the strong subadditivity of von Neumann entropy or Lieb's concavity theorem.

Quantum random access codes were proposed by Ambainis, Nayak, Ta-Shma, and Vazirani [8]; they proved a somewhat weaker limitation on quantum random access codes than what is established by Corollary 5.58, which

was proved by Nayak [162] a short time later. (The two previously referenced papers appeared in conference proceedings, and were consolidated as a journal paper [9].) Nayak's theorem, as stated in Theorem 5.56, follows from the proof of a closely related theorem that appears in Nayak's PhD thesis [161].

Chapter 6

Bipartite entanglement

Entanglement is a fundamental concept in quantum information theory, considered by many to be a quintessential characteristic that distinguishes quantum systems from their classical counterparts. Informally speaking, a state of a collection of registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ is said to be *entangled* when it is not possible to specify the correlations that exist among the registers in classical terms. When it is possible to describe these correlations in classical terms, the registers are said to be in a *separable* state. Entanglement among two or more registers is therefore synonymous with a lack of separability in their state.

This chapter introduces notions associated with bipartite entanglement, in which correlations between precisely two registers (or two collections of registers) are considered. Topics to be discussed include the property of separability, which is applicable not only to states but also to channels and measurements; aspects of entanglement manipulation and quantification; and a discussion of operational phenomena associated with entanglement, including teleportation, dense coding, and non-classical correlations among measurements on separated systems.

6.1 Separability

This section introduces the notion of separability, which is applicable to states, channels, and measurements on bipartite systems. It is possible to define a multipartite variant of this concepts, but only bipartite separability is considered in this book.

6.1.1 Separable operators and states

The property of separability for operators acting on bipartite tensor product spaces is defined as follows.

Definition 6.1. For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, the set $\mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ is defined as the set containing all positive semidefinite operators $R \in \mathrm{Pos}(\mathcal{X}\otimes\mathcal{Y})$ for which there exists an alphabet Σ and two collections of positive semidefinite operators,

$$\{P_a \,:\, a\in\Sigma\} \subset \mathrm{Pos}(\mathcal{X}) \quad \text{and} \quad \{Q_a \,:\, a\in\Sigma\} \subset \mathrm{Pos}(\mathcal{Y}), \tag{6.1}$$

such that

$$R = \sum_{a\in\Sigma} P_a \otimes Q_a. \tag{6.2}$$

Elements of the set $\mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ are called *separable operators*.

Remark 6.2. As the previous definition reflects, separability is defined with respect to a particular tensor product structure of the underlying complex Euclidean space of a given operator. When the term *separable operator* is used, one must therefore make this tensor product structure known (if it is not implicit). For instance, an operator $R \in \mathrm{Pos}(\mathcal{X}\otimes\mathcal{Y}\otimes\mathcal{Z})$ may be an element of $\mathrm{Sep}(\mathcal{X}:\mathcal{Y}\otimes\mathcal{Z})$ but not $\mathrm{Sep}(\mathcal{X}\otimes\mathcal{Y}:\mathcal{Z})$.

By restricting the definition above to density operators, one obtains a definition of *separable states*.

Definition 6.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces. One defines

$$\mathrm{SepD}(\mathcal{X}:\mathcal{Y}) = \mathrm{Sep}(\mathcal{X}:\mathcal{Y}) \cap \mathrm{D}(\mathcal{X}\otimes\mathcal{Y}). \tag{6.3}$$

Elements of the set $\mathrm{SepD}(\mathcal{X}:\mathcal{Y})$ are referred to as *separable states* (or *separable density operators*).

Convex properties of separable operators and states

The sets $\mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ and $\mathrm{SepD}(\mathcal{X}:\mathcal{Y})$ possess various properties relating to convexity, a few of which will now be observed.

Proposition 6.4. *For every choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, the set $\mathrm{SepD}(\mathcal{X}:\mathcal{Y})$ is convex, and the set $\mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ is a convex cone.*

Proof. It will first be proved that $\mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ is a convex cone. It suffices to prove that $\mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ is closed under addition as well as multiplication by any nonnegative real number. To this end, assume that $R_0, R_1 \in \mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ are separable operators and $\lambda \geq 0$ is a nonnegative real number. One may write

$$R_0 = \sum_{a\in\Sigma_0} P_a \otimes Q_a \quad \text{and} \quad R_1 = \sum_{a\in\Sigma_1} P_a \otimes Q_a \tag{6.4}$$

for disjoint alphabets Σ_0 and Σ_1, where

$$\{P_a \,:\, a \in \Sigma_0 \cup \Sigma_1\} \subset \mathrm{Pos}(\mathcal{X}) \;\text{ and }\; \{Q_a \,:\, a \in \Sigma_0 \cup \Sigma_1\} \subset \mathrm{Pos}(\mathcal{Y}) \tag{6.5}$$

are collections of positive semidefinite operators. It holds that

$$R_0 + R_1 = \sum_{a\in\Sigma_0\cup\Sigma_1} P_a \otimes Q_a, \tag{6.6}$$

and therefore $R_0 + R_1 \in \mathrm{Sep}(\mathcal{X}:\mathcal{Y})$. Moreover, it holds that

$$\lambda R_0 = \sum_{a\in\Sigma_0} (\lambda P_a) \otimes Q_a. \tag{6.7}$$

As $\lambda P \in \mathrm{Pos}(\mathcal{X})$ for every positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X})$, it follows that $\lambda R_0 \in \mathrm{Sep}(\mathcal{X}:\mathcal{Y})$.

The fact that $\mathrm{SepD}(\mathcal{X}:\mathcal{Y})$ is convex follows from the fact that it is equal to the intersection of two convex sets, $\mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ and $\mathrm{D}(\mathcal{X}\otimes\mathcal{Y})$. □

The next proposition, when combined with the previous one, implies that $\mathrm{Sep}(\mathcal{X}:\mathcal{Y})$ is equal to the cone generated by $\mathrm{SepD}(\mathcal{X}:\mathcal{Y})$.

Proposition 6.5. *Let $\mathcal{Z}$ be a complex Euclidean space, let $\mathcal{A} \subseteq \mathrm{Pos}(\mathcal{Z})$ be a cone, and assume that $\mathcal{B} = \mathcal{A} \cap \mathrm{D}(\mathcal{Z})$ is nonempty. It holds that*

$$\mathcal{A} = \big\{\lambda\rho \,:\, \lambda \geq 0,\ \rho \in \mathcal{B}\big\}. \tag{6.8}$$

Proof. One may write

$$\mathrm{cone}(\mathcal{B}) = \big\{\lambda\rho \,:\, \lambda \geq 0,\ \rho \in \mathcal{B}\big\} \tag{6.9}$$

for brevity. Suppose first that $\rho \in \mathcal{B}$ and $\lambda \geq 0$. It follows that $\lambda\rho \in \mathcal{A}$ by virtue of the fact that $\mathcal{B} \subseteq \mathcal{A}$ and $\mathcal{A}$ is a cone. Therefore $\mathrm{cone}(\mathcal{B}) \subseteq \mathcal{A}$.

Now suppose that $P \in \mathcal{A}$. If $P = 0$, then one has that $P = \lambda\rho$ for $\lambda = 0$ and $\rho \in \mathcal{B}$ being chosen arbitrarily. If $P \neq 0$, then consider the density operator $\rho = P/\operatorname{Tr}(P)$. It holds that $\rho \in \mathcal{A}$ because $1/\operatorname{Tr}(P) > 0$ and $\mathcal{A}$ is a cone, and therefore $\rho \in \mathcal{B}$. As $P = \lambda\rho$ for $\lambda = \operatorname{Tr}(P) > 0$, it follows that $P \in \mathrm{cone}(\mathcal{B})$. Therefore, $\mathcal{A} \subseteq \mathrm{cone}(\mathcal{B})$, which completes the proof. □

Two equivalent ways of specifying separable states are provided by the next proposition, which is a straightforward consequence of the spectral theorem.

Proposition 6.6. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\xi \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be a density operator. The following statements are equivalent:*

1. $\xi \in \mathrm{SepD}(\mathcal{X} : \mathcal{Y})$.
2. *There exists an alphabet Σ, two collections of states $\{\rho_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X})$ and $\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y})$, and a probability vector $p \in \mathcal{P}(\Sigma)$, such that*

$$\xi = \sum_{a \in \Sigma} p(a)\, \rho_a \otimes \sigma_a. \tag{6.10}$$

3. *There exists an alphabet Σ, two collections of unit vectors $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ and $\{y_a : a \in \Sigma\} \subset \mathcal{Y}$, and a probability vector $p \in \mathcal{P}(\Sigma)$, such that*

$$\xi = \sum_{a \in \Sigma} p(a)\, x_a x_a^* \otimes y_a y_a^*. \tag{6.11}$$

Proof. The third statement trivially implies the second, and it is immediate that the second statement implies the first, as $\mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ is convex and $\rho_a \otimes \sigma_a \in \mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ for each $a \in \Sigma$. It remains to prove that the first statement implies the third.

Let $\xi \in \mathrm{SepD}(\mathcal{X} : \mathcal{Y})$. As $\xi \in \mathrm{Sep}(\mathcal{X} : \mathcal{Y})$, one may write

$$\xi = \sum_{b \in \Gamma} P_b \otimes Q_b \tag{6.12}$$

for some choice of an alphabet Γ and collections $\{P_b : b \in \Gamma\} \subset \mathrm{Pos}(\mathcal{X})$ and $\{Q_b : b \in \Gamma\} \subset \mathrm{Pos}(\mathcal{Y})$ of positive semidefinite operators. Let $n = \dim(\mathcal{X})$, let $m = \dim(\mathcal{X})$, and consider spectral decompositions of these operators as follows:

$$P_b = \sum_{j=1}^{n} \lambda_j(P_b) u_{b,j} u_{b,j}^* \quad \text{and} \quad Q_b = \sum_{k=1}^{m} \lambda_k(Q_b) v_{b,k} v_{b,k}^*, \tag{6.13}$$

for each $b \in \Gamma$. Define $\Sigma = \Gamma \times \{1, \ldots, n\} \times \{1, \ldots, m\}$, and define

$$\begin{aligned} p((b,j,k)) &= \lambda_j(P_b)\lambda_k(Q_b), \\ x_{(b,j,k)} &= u_{b,j}, \\ y_{(b,j,k)} &= v_{b,k}, \end{aligned} \tag{6.14}$$

for every $(b, j, k) \in \Sigma$. A straightforward computation reveals that

$$\sum_{a \in \Sigma} p(a)\, x_a x_a^* \otimes y_a y_a^* = \sum_{b \in \Gamma} P_b \otimes Q_b = \xi. \tag{6.15}$$

Moreover, each value $p(a)$ is nonnegative, and because

$$\sum_{a \in \Sigma} p(a) = \mathrm{Tr}(\xi) = 1, \tag{6.16}$$

it follows that that p is a probability vector. It has therefore been proved that statement 1 implies statement 3. □

By the equivalence of the first and second statements in the previous proposition, it holds that a given separable state $\xi \in \mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ represents a classical probability distribution over independent quantum states of a pair of registers (X, Y); and in this sense the possible states of the registers X and Y, when considered in isolation, are classically correlated.

For a separable state $\xi \in \mathrm{SepD}(\mathcal{X} : \mathcal{Y})$, the expression (6.11) is generally not unique—there may be many inequivalent ways that ξ can be expressed in this form. It is important to observe that an expression of this form cannot necessarily be obtained directly from a spectral decomposition of ξ. Indeed, for some choices of $\xi \in \mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ it may hold that every expression of ξ in the form (6.11) requires that Σ has cardinality strictly larger than $\mathrm{rank}(\xi)$. An upper bound on the size of the alphabet Σ required for an expression of the form (6.11) to exist may, however, be obtained from Carathéodory's theorem (Theorem 1.9).

Proposition 6.7. *Let $\xi \in \mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. There exists an alphabet Σ such that $|\Sigma| \leq \mathrm{rank}(\xi)^2$, two collections of unit vectors $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ and $\{y_a : a \in \Sigma\} \subset \mathcal{Y}$, and a probability vector $p \in \mathcal{P}(\Sigma)$ such that*

$$\xi = \sum_{a \in \Sigma} p(a)\, x_a x_a^* \otimes y_a y_a^*. \tag{6.17}$$

Proof. By Proposition 6.6 it holds that

$$\mathrm{SepD}(\mathcal{X} : \mathcal{Y}) = \mathrm{conv}\bigl\{ x x^* \otimes y y^* : x \in \mathcal{S}(\mathcal{X}),\, y \in \mathcal{S}(\mathcal{Y}) \bigr\}, \tag{6.18}$$

from which it follows that ξ is contained in the set

$$\mathrm{conv}\bigl\{ x x^* \otimes y y^* : x \in \mathcal{S}(\mathcal{X}),\, y \in \mathcal{S}(\mathcal{Y}),\, \mathrm{im}(x x^* \otimes y y^*) \subseteq \mathrm{im}(\xi) \bigr\}. \tag{6.19}$$

Every density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ satisfying $\mathrm{im}(\rho) \subseteq \mathrm{im}(\xi)$ is contained in the real affine subspace

$$\big\{H \in \mathrm{Herm}(\mathcal{X} \otimes \mathcal{Y}) \,:\, \mathrm{im}(H) \subseteq \mathrm{im}(\xi),\, \mathrm{Tr}(H) = 1\big\} \tag{6.20}$$

of dimension $\mathrm{rank}(\xi)^2 - 1$, and therefore the proposition follows directly from Carathéodory's theorem. □

By combining the previous proposition with Proposition 6.5, one obtains the following corollary.

Corollary 6.8. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $R \in \mathrm{Sep}(\mathcal{X} : \mathcal{Y})$ be a nonzero operator. There exists an alphabet Σ such that $|\Sigma| \leq \mathrm{rank}(R)^2$, along with two collections of vectors $\{x_a \,:\, a \in \Sigma\} \subset \mathcal{X}$ and $\{y_a \,:\, a \in \Sigma\} \subset \mathcal{Y}$, such that*

$$R = \sum_{a \in \Sigma} x_a x_a^* \otimes y_a y_a^*. \tag{6.21}$$

The last observation to be made about separable operators and states in this subsection is the following proposition, which establishes a basic topological property of the sets $\mathrm{Sep}(\mathcal{X} : \mathcal{Y})$ and $\mathrm{SepD}(\mathcal{X} : \mathcal{Y})$.

Proposition 6.9. *For every choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, the set $\mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ is compact and the set $\mathrm{Sep}(\mathcal{X} : \mathcal{Y})$ is closed.*

Proof. The unit spheres $\mathcal{S}(\mathcal{X})$ and $\mathcal{S}(\mathcal{Y})$ are compact, which implies that their Cartesian product $\mathcal{S}(\mathcal{X}) \times \mathcal{S}(\mathcal{Y})$ is also compact. The function

$$\phi : \mathcal{S}(\mathcal{X}) \times \mathcal{S}(\mathcal{Y}) \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}) : (x, y) \mapsto xx^* \otimes yy^* \tag{6.22}$$

is continuous, and therefore the set

$$\phi(\mathcal{S}(\mathcal{X}) \times \mathcal{S}(\mathcal{Y})) = \big\{xx^* \otimes yy^* \,:\, x \in \mathcal{S}(\mathcal{X}),\, y \in \mathcal{S}(\mathcal{Y})\big\} \tag{6.23}$$

is compact. Because the convex hull of a compact set is necessarily compact, it follows that $\mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ is compact.

As $\mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ is compact, and does not include 0, the cone it generates must be closed, and therefore $\mathrm{Sep}(\mathcal{X} : \mathcal{Y})$ is closed. □

The Horodecki criterion

The next theorem provides an alternative characterization of separability, demonstrating that the property of separability for operators has a close connection with the property of positivity for maps.

Theorem 6.10 (Horodecki criterion)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $R \in \operatorname{Pos}(\mathcal{X} \otimes \mathcal{Y})$ be a positive semidefinite operator. The following three statements are equivalent:*

1. $R \in \operatorname{Sep}(\mathcal{X} : \mathcal{Y})$.
2. *For every complex Euclidean space $\mathcal{Z}$ and every positive map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Z})$ it holds that*
$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(R) \in \operatorname{Pos}(\mathcal{Z} \otimes \mathcal{Y}). \tag{6.24}$$
3. *For every positive and unital map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$, it holds that*
$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(R) \in \operatorname{Pos}(\mathcal{Y} \otimes \mathcal{Y}). \tag{6.25}$$

Proof. Suppose first that $R \in \operatorname{Sep}(\mathcal{X} : \mathcal{Y})$, so that

$$R = \sum_{a \in \Sigma} P_a \otimes Q_a \tag{6.26}$$

for some choice of an alphabet Σ and collections $\{P_a : a \in \Sigma\} \subset \operatorname{Pos}(\mathcal{X})$ and $\{Q_a : a \in \Sigma\} \subset \operatorname{Pos}(\mathcal{Y})$. For every complex Euclidean space $\mathcal{Z}$ and every positive map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Z})$ it holds that

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(R) = \sum_{a \in \Sigma} \Phi(P_a) \otimes Q_a \in \operatorname{Pos}(\mathcal{Z} \otimes \mathcal{Y}), \tag{6.27}$$

by virtue of the fact that $\Phi(P_a)$ is a positive semidefinite operator for each $a \in \Sigma$. Statement 1 therefore implies statement 2.

Statement 2 trivially implies statement 3.

Finally, the fact that statement 3 implies statement 1 will be proved in the contrapositive form. To this end, assume $R \in \operatorname{Pos}(\mathcal{X} \otimes \mathcal{Y})$ is not a separable operator. As $\operatorname{Sep}(\mathcal{X} : \mathcal{Y})$ is a closed, convex cone within the real vector space $\operatorname{Herm}(\mathcal{X} \otimes \mathcal{Y})$, the hyperplane separation theorem (Theorem 1.11) implies that there must exist a Hermitian operator $H \in \operatorname{Herm}(\mathcal{X} \otimes \mathcal{Y})$ such that

$\langle H, R\rangle < 0$ and $\langle H, S\rangle \geq 0$ for every $S \in \mathrm{Sep}(\mathcal{X} : \mathcal{Y})$. The operator H will be used to define a positive and unital map $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ for which

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(R) \notin \mathrm{Pos}(\mathcal{Y} \otimes \mathcal{Y}). \tag{6.28}$$

First, let $\Psi \in \mathrm{T}(\mathcal{Y}, \mathcal{X})$ be the unique map for which $J(\Psi) = H$, choose $\varepsilon > 0$ to be a sufficiently small positive real number so that the inequality

$$\langle H, R\rangle + \varepsilon \,\mathrm{Tr}(R) < 0 \tag{6.29}$$

is satisfied, and define $\Xi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ as

$$\Xi(X) = \Psi^*(X) + \varepsilon \,\mathrm{Tr}(X)\mathbb{1}_{\mathcal{Y}} \tag{6.30}$$

for every $X \in \mathrm{L}(\mathcal{X})$. For arbitrarily chosen positive semidefinite operators $P \in \mathrm{Pos}(\mathcal{X})$ and $Q \in \mathrm{Pos}(\mathcal{Y})$, it is the case that

$$P \otimes \overline{Q} \in \mathrm{Sep}(\mathcal{X} : \mathcal{Y}), \tag{6.31}$$

and therefore

$$0 \leq \big\langle H, P \otimes \overline{Q}\big\rangle = \big\langle P \otimes \overline{Q}, J(\Psi)\big\rangle = \big\langle P, \Psi(Q)\big\rangle. \tag{6.32}$$

The fact that this inequality holds for every choice of $P \in \mathrm{Pos}(\mathcal{X})$ and $Q \in \mathrm{Pos}(\mathcal{Y})$ implies that $\Psi(Q) \in \mathrm{Pos}(\mathcal{X})$ for every choice of $Q \in \mathrm{Pos}(\mathcal{Y})$, and therefore Ψ is a positive map. It follows from Proposition 2.17 that Ψ^* is a positive map as well. For every nonzero positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X})$, the operator $\Xi(P)$ is therefore equal to a positive semidefinite operator $\Psi^*(P)$ plus a positive multiple of the identity operator.

Now let $A = \Xi(\mathbb{1}_{\mathcal{X}})$, which is necessarily a positive definite operator, and define $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ as

$$\Phi(X) = A^{-\frac{1}{2}}\Xi(X)A^{-\frac{1}{2}} \tag{6.33}$$

for every $X \in \mathrm{L}(\mathcal{X})$. It remains to verify that Φ is indeed a positive and unital map for which (6.28) holds. The positivity of Φ follows from the fact that Ξ is positive, and it holds that

$$\Phi(\mathbb{1}_{\mathcal{X}}) = A^{-\frac{1}{2}}\Xi(\mathbb{1}_{\mathcal{X}})A^{-\frac{1}{2}} = A^{-\frac{1}{2}}AA^{-\frac{1}{2}} = \mathbb{1}_{\mathcal{Y}}, \tag{6.34}$$

establishing that Φ is unital. Finally, through the following computation one may verify that the operator $\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(R)$ is not positive semidefinite:

$$\begin{aligned}
&\Big\langle \operatorname{vec}\big(\sqrt{A}\big) \operatorname{vec}\big(\sqrt{A}\big)^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(R) \Big\rangle \\
&\qquad = \Big\langle \operatorname{vec}\big(\mathbb{1}_{\mathcal{Y}}\big) \operatorname{vec}\big(\mathbb{1}_{\mathcal{Y}}\big)^*, \big(\Xi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(R) \Big\rangle \\
&\qquad = \big\langle J(\Xi^*), R \big\rangle \\
&\qquad = \big\langle J(\Psi) + \varepsilon \mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}}, R \big\rangle \\
&\qquad = \langle H, R \rangle + \varepsilon \operatorname{Tr}(R) \\
&\qquad < 0.
\end{aligned} \tag{6.35}$$

This completes the proof. □

One immediate application of Theorem 6.10 is that it provides a method for proving that certain positive semidefinite operators are not separable. The following example demonstrates this method for two families of states known as *Werner states* and *isotropic states*.

Example 6.11. Let Σ be an alphabet, and let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces of the form $\mathcal{X} = \mathbb{C}^{\Sigma}$ and $\mathcal{Y} = \mathbb{C}^{\Sigma}$. The *swap operator* $W \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ is the unique operator satisfying

$$W(x \otimes y) = y \otimes x \tag{6.36}$$

for all vectors $x, y \in \mathbb{C}^{\Sigma}$. Equivalently, this operator is given by

$$W = \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{b,a}. \tag{6.37}$$

The operator W is both unitary and Heritian, having eigenvalues 1 and -1. The eigenspace of W corresponding to the eigenvalue 1 is spanned by the orthonormal collection

$$\left\{ \frac{e_a \otimes e_b + e_b \otimes e_a}{\sqrt{2}} \,:\, a, b \in \Sigma,\, a \neq b \right\} \cup \big\{ e_a \otimes e_a \,:\, a \in \Sigma \big\}, \tag{6.38}$$

while the eigenspace corresponding to the eigenvalue -1 is spanned by the orthonormal collection

$$\left\{ \frac{e_a \otimes e_b - e_b \otimes e_a}{\sqrt{2}} \,:\, a, b \in \Sigma,\, a \neq b \right\}. \tag{6.39}$$

Let $n = |\Sigma|$, and define projection operators Δ_0, Δ_1, Π_0, $\Pi_1 \in \mathrm{Proj}(\mathcal{X} \otimes \mathcal{Y})$ as follows:

$$\Delta_0 = \frac{1}{n} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b}\,, \quad \Pi_0 = \frac{1}{2}\mathbb{1} \otimes \mathbb{1} + \frac{1}{2}W\,, \tag{6.40}$$

$$\Delta_1 = \mathbb{1} \otimes \mathbb{1} - \Delta_0\,, \qquad \Pi_1 = \mathbb{1} \otimes \mathbb{1} - \Pi_0\,. \tag{6.41}$$

That these operators are indeed projection operators follows from the fact that they are Hermitian and square to themselves. Alternatively, one may observe that $\Delta_0 = uu^*$ is the projection onto the one-dimensional subspace of $\mathcal{X} \otimes \mathcal{Y}$ spanned by the unit vector

$$u = \frac{1}{\sqrt{n}} \sum_{a \in \Sigma} e_a \otimes e_a, \tag{6.42}$$

Δ_1 is the projection onto the orthogonal complement of this subspace, and Π_0 and Π_1 are the projection operators onto the subspaces spanned by the collections (6.38) and (6.39), respectively. (The images of Π_0 and Π_1 are also known as the *symmetric* and *antisymmetric* subspaces of $\mathbb{C}^\Sigma \otimes \mathbb{C}^\Sigma$, and are considered in greater detail and generality in Chapter 7.) It holds that

$$\begin{aligned} &\mathrm{rank}(\Delta_0) = 1, && \mathrm{rank}(\Pi_0) = \binom{n+1}{2}, \\ &\mathrm{rank}(\Delta_1) = n^2 - 1, && \mathrm{rank}(\Pi_1) = \binom{n}{2}. \end{aligned} \tag{6.43}$$

States of the form

$$\lambda \Delta_0 + (1 - \lambda) \frac{\Delta_1}{n^2 - 1} \tag{6.44}$$

are known as *isotropic states*, and states of the form

$$\lambda \frac{\Pi_0}{\binom{n+1}{2}} + (1 - \lambda) \frac{\Pi_1}{\binom{n}{2}} \tag{6.45}$$

are known as *Werner states* (for $\lambda \in [0, 1]$ in both cases).

Now, let $\mathrm{T} \in \mathrm{T}(\mathcal{X})$ denote the transpose mapping, defined by the action $\mathrm{T}(X) = X^{\mathsf{T}}$ for all $X \in \mathrm{L}(\mathcal{X})$. The mapping T is a positive map. Using the observation that

$$(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\Delta_0) = \frac{1}{n} W, \tag{6.46}$$

which may be verified directly, as well as $\mathrm{T}(\mathbb{1}_{\mathcal{X}}) = \mathbb{1}_{\mathcal{X}}$ and $\mathrm{T}^2 = \mathbb{1}_{\mathrm{L}(\mathcal{X})}$, the following relations may be obtained:

$$(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\Delta_0) = \frac{1}{n}\Pi_0 - \frac{1}{n}\Pi_1\,, \tag{6.47}$$

$$(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\Delta_1) = \frac{n-1}{n}\Pi_0 + \frac{n+1}{n}\Pi_1\,, \tag{6.48}$$

$$(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\Pi_0) = \frac{n+1}{2}\Delta_0 + \frac{1}{2}\Delta_1\,, \tag{6.49}$$

$$(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\Pi_1) = -\frac{n-1}{2}\Delta_0 + \frac{1}{2}\Delta_1\,. \tag{6.50}$$

For $\lambda \in [0,1]$, the equations

$$\begin{aligned}(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})&\left(\lambda\Delta_0 + (1-\lambda)\frac{\Delta_1}{n^2-1}\right)\\ &= \left(\frac{1+\lambda n}{2}\right)\frac{\Pi_0}{\binom{n+1}{2}} + \left(\frac{1-\lambda n}{2}\right)\frac{\Pi_1}{\binom{n}{2}}\end{aligned} \tag{6.51}$$

and

$$\begin{aligned}(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})&\left(\lambda\frac{\Pi_0}{\binom{n+1}{2}} + (1-\lambda)\frac{\Pi_1}{\binom{n}{2}}\right)\\ &= \left(\frac{2\lambda-1}{n}\right)\Delta_0 + \left(1 - \frac{2\lambda-1}{n}\right)\frac{\Delta_1}{n^2-1}\end{aligned} \tag{6.52}$$

are implied. It therefore holds that the isotropic state (6.44) is entangled (i.e., not separable) for $\lambda \in (1/n, 1]$, while the Werner state (6.45) is entangled for $\lambda \in [0, 1/2)$.[1]

A separable neighborhood of the identity operator

By means of the Horodecki criterion (Theorem 6.10), it may be proved that there exists a neighborhood of the identity operator $\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}}$, for any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, in which *every* positive semidefinite operator is separable. Consequently, every density operator $\mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ that is sufficiently close to the completely mixed state is separable. In order to

[1] It does indeed hold that the isotropic state (6.44) is separable for $\lambda \in [0, 1/n]$ and the Werner state (6.45) is separable for $\lambda \in [1/2, 1]$. These facts are proved in Chapter 7 (q.v. Example 7.26).

prove this fact, which is stated in more precise terms in Theorem 6.14 below, the following lemma will be used.

Lemma 6.12. *Let Σ be an alphabet, let $\mathcal{X}$ and $\mathcal{Y} = \mathbb{C}^{\Sigma}$ be complex Euclidean spaces, and let $\{A_{a,b} : a, b \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$ be a collection of operators. For $A \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ being the operator defined as*

$$A = \sum_{a,b \in \Sigma} A_{a,b} \otimes E_{a,b}\,, \tag{6.53}$$

one has that

$$\|A\|^2 \leq \sum_{a,b \in \Sigma} \|A_{a,b}\|^2. \tag{6.54}$$

Proof. For each $a \in \Sigma$, define an operator $B_a \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ as

$$B_a = \sum_{b \in \Sigma} A_{a,b} \otimes E_{a,b}. \tag{6.55}$$

By expanding the product $B_a B_a^*$ and applying the triangle inequality, the multiplicativity of the spectral norm under tensor products, and the spectral norm identity (1.172), one finds that

$$\|B_a B_a^*\| = \left\| \sum_{b \in \Sigma} A_{a,b} A_{a,b}^* \otimes E_{a,a} \right\| \leq \sum_{b \in \Sigma} \|A_{a,b} A_{a,b}^*\| = \sum_{b \in \Sigma} \|A_{a,b}\|^2. \tag{6.56}$$

It holds that

$$A^* A = \sum_{a \in \Sigma} B_a^* B_a, \tag{6.57}$$

and therefore, by (6.56) together with the triangle inequality and spectral norm identity,

$$\|A\|^2 = \|A^* A\| \leq \sum_{a \in \Sigma} \|B_a^* B_a\| \leq \sum_{a,b \in \Sigma} \|A_{a,b}\|^2, \tag{6.58}$$

as required. □

In addition, the following theorem (which is equivalent to Theorem 3.42) will be needed.

Theorem 6.13. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be a positive and unital map. It holds that*

$$\|\Phi(X)\| \leq \|X\| \tag{6.59}$$

for every operator $X \in \mathrm{L}(\mathcal{X})$.

Proof. By the assumption that Φ is positive and unital, Proposition 2.17 and Theorem 2.26 imply that Φ^* is positive and trace-preserving. For operators $X \in \mathrm{L}(\mathcal{X})$ and $Y \in \mathrm{L}(\mathcal{Y})$, one therefore has

$$\begin{aligned} \big|\langle Y, \Phi(X)\rangle\big| = \big|\langle \Phi^*(Y), X\rangle\big| &\leq \|X\|\big\|\Phi^*(Y)\big\|_1 \\ &\leq \|X\|\|Y\|_1\|\Phi^*\|_1 = \|X\|\|Y\|_1, \end{aligned} \tag{6.60}$$

where the final equality follows by Corollary 3.43 (to Theorem 3.42). By maximizing over all operators $Y \in \mathrm{L}(\mathcal{Y})$ that satisfy $\|Y\|_1 \leq 1$, one finds that $\|\Phi(X)\| \leq \|X\|$ for every $X \in \mathrm{L}(\mathcal{X})$, as required. □

Theorem 6.14. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $H \in \mathrm{Herm}(\mathcal{X} \otimes \mathcal{Y})$ be a Hermitian operator, and assume that $\|H\|_2 \leq 1$. It holds that*

$$\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - H \in \mathrm{Sep}(\mathcal{X} : \mathcal{Y}). \tag{6.61}$$

Proof. Let $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ be an arbitrarily chosen positive and unital map. Let Σ be the alphabet for which $\mathcal{Y} = \mathbb{C}^{\Sigma}$, and write

$$H = \sum_{a,b \in \Sigma} H_{a,b} \otimes E_{a,b}. \tag{6.62}$$

It holds that

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(H) = \sum_{a,b \in \Sigma} \Phi(H_{a,b}) \otimes E_{a,b}, \tag{6.63}$$

and therefore

$$\begin{aligned} \big\|\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(H)\big\|^2 &\leq \sum_{a,b \in \Sigma} \|\Phi(H_{a,b})\|^2 \\ &\leq \sum_{a,b \in \Sigma} \|H_{a,b}\|^2 \leq \sum_{a,b \in \Sigma} \|H_{a,b}\|_2^2 = \|H\|_2^2 \leq 1. \end{aligned} \tag{6.64}$$

(The first inequality is implied by Lemma 6.12, and the second inequality is implied by Theorem 6.13.) The positivity of Φ implies that $(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(H)$ is Hermitian, and therefore $(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(H) \leq \mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}}$. It follows that

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - H) = \mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(H) \geq 0. \tag{6.65}$$

Because (6.65) holds for all positive and unital maps Φ, one concludes from Theorem 6.10 that $\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - H$ is separable. □

Bipartite operator entanglement rank

Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and consider the collection of all positive semidefinite operators $R \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ for which there exists an alphabet Σ and a collection of operators $\{X_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y}, \mathcal{X})$ such that

$$R = \sum_{a \in \Sigma} \mathrm{vec}(X_a)\,\mathrm{vec}(X_a)^* \tag{6.66}$$

and $\mathrm{rank}(X_a) \leq 1$ for each $a \in \Sigma$. It holds that an operator $X \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ has rank at most 1 if and only if there exist vectors $u \in \mathcal{X}$ and $v \in \mathcal{Y}$ such that $\mathrm{vec}(X) = u \otimes v$, and from this observation it follows that the collection of operators R just described coincides with $\mathrm{Sep}(\mathcal{X} : \mathcal{Y})$.

It is useful to generalize this notion, allowing for arbitrary upper-bounds on the rank of the operators $\{X_a : a \in \Sigma\}$, along the lines of the following definition.

Definition 6.15. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $r \geq 1$ be a positive integer. The set $\mathrm{Ent}_r(\mathcal{X} : \mathcal{Y})$ is defined to be the set of all operators $R \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ for which there exists an alphabet Σ and a collection of operators

$$\{X_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y}, \mathcal{X}) \tag{6.67}$$

satisfying $\mathrm{rank}(X_a) \leq r$ for each $a \in \Sigma$, such that

$$R = \sum_{a \in \Sigma} \mathrm{vec}(X_a)\,\mathrm{vec}(X_a)^*. \tag{6.68}$$

An element $R \in \mathrm{Ent}_r(\mathcal{X} : \mathcal{Y})$ is said to have *entanglement rank* bounded by r. The *entanglement rank* of $R \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$, with respect to the bipartition between $\mathcal{X}$ and $\mathcal{Y}$, is the minimum value of $r \geq 1$ such that $R \in \mathrm{Ent}_r(\mathcal{X} : \mathcal{Y})$.

As indicated above, it holds that

$$\mathrm{Sep}(\mathcal{X} : \mathcal{Y}) = \mathrm{Ent}_1(\mathcal{X} : \mathcal{Y}), \tag{6.69}$$

and from Definition 6.15 it is immediate that

$$\mathrm{Ent}_{r-1}(\mathcal{X} : \mathcal{Y}) \subseteq \mathrm{Ent}_r(\mathcal{X} : \mathcal{Y}) \tag{6.70}$$

for every integer $r \geq 2$.

The containment (6.70) is proper, provided $r \leq \min\{\dim(\mathcal{X}), \dim(\mathcal{Y})\}$. To see that this is so, consider any operator $Y \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ having rank equal to r, and suppose that

$$\operatorname{vec}(Y)\operatorname{vec}(Y)^* = \sum_{a\in\Sigma} \operatorname{vec}(X_a)\operatorname{vec}(X_a)^* \tag{6.71}$$

for some collection of operators $\{X_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y}, \mathcal{X})$. As the operator represented by this equation has rank equal to 1, it must hold that $X_a = \alpha_a Y$ for each $a \in \Sigma$, for $\{\alpha_a : a \in \Sigma\}$ being a collection of complex numbers satisfying

$$\sum_{a\in\Sigma} |\alpha_a|^2 = 1. \tag{6.72}$$

It is therefore not possible that (6.71) holds when each operator X_a has rank strictly smaller than r, and therefore

$$\operatorname{vec}(Y)\operatorname{vec}(Y)^* \notin \mathrm{Ent}_{r-1}(\mathcal{X} : \mathcal{Y}). \tag{6.73}$$

It is immediate, on the other hand, that $\operatorname{vec}(Y)\operatorname{vec}(Y)^* \in \mathrm{Ent}_r(\mathcal{X} : \mathcal{Y})$.

Finally, one may observe that

$$\mathrm{Ent}_n(\mathcal{X} : \mathcal{Y}) = \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}) \tag{6.74}$$

for $n \geq \min\{\dim(\mathcal{X}), \dim(\mathcal{Y})\}$, as every operator $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ has rank bounded by n in this case.

The following simple proposition concerning entanglement rank will be useful in the subsequent sections of this chapter.

Proposition 6.16. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $Y \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$, and assume that $\|Y\| \leq 1$. For every positive integer r and every operator*

$$P \in \mathrm{Ent}_r(\mathcal{X} : \mathcal{Y}) \tag{6.75}$$

having entanglement rank bounded by r, it holds that

$$\big\langle \operatorname{vec}(Y)\operatorname{vec}(Y)^*, P \big\rangle \leq r \operatorname{Tr}(P). \tag{6.76}$$

Proof. Under the assumption that P has entanglement rank bounded by r, one may write

$$P = \sum_{a\in\Sigma} \operatorname{vec}(X_a)\operatorname{vec}(X_a)^* \tag{6.77}$$

for some alphabet Σ and a collection of operators $\{X_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y}, \mathcal{X})$ for which $\operatorname{rank}(X_a) \leq r$ for every $a \in \Sigma$. For every operator $X \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$, one has

$$\big|\langle Y, X\rangle\big|^2 \leq \|X\|_1^2 \leq \operatorname{rank}(X)\,\|X\|_2^2, \tag{6.78}$$

so that evaluating the inner product in the statement of the proposition yields

$$\begin{aligned}\big\langle \operatorname{vec}(Y)\operatorname{vec}(Y)^*, P\big\rangle &= \sum_{a\in\Sigma} \big|\langle Y, X_a\rangle\big|^2 \\ &\leq \sum_{a\in\Sigma} \operatorname{rank}(X_a)\big\|X_a\big\|_2^2 \leq r\sum_{a\in\Sigma}\big\|X_a\big\|_2^2 = r\operatorname{Tr}(P),\end{aligned} \tag{6.79}$$

as required. □

Example 6.17. Let Σ be an alphabet, let $n = |\Sigma|$, let $\mathcal{X} = \mathbb{C}^\Sigma$ and $\mathcal{Y} = \mathbb{C}^\Sigma$, and define a density operator $\tau \in \mathrm{D}(\mathcal{X}\otimes\mathcal{Y})$ as

$$\tau = \frac{1}{n}\sum_{a,b\in\Sigma} E_{a,b}\otimes E_{a,b}. \tag{6.80}$$

The density operator τ, which coincides with the isotropic state Δ_0 defined in Example 6.11, is the canonical example of a maximally entangled state with respect to the spaces $\mathcal{X}$ and $\mathcal{Y}$. One may observe that

$$\tau = \frac{1}{n}\operatorname{vec}(\mathbb{1})\operatorname{vec}(\mathbb{1})^* \tag{6.81}$$

for $\mathbb{1}$ denoting the identity operator on $\mathbb{C}^\Sigma$, which may be viewed as an element of the set $\mathrm{L}(\mathcal{Y}, \mathcal{X})$ in the most straightforward way.

For every positive integer r and every density operator

$$\rho \in \mathrm{D}(\mathcal{X}\otimes\mathcal{Y}) \cap \mathrm{Ent}_r(\mathcal{X}:\mathcal{Y}) \tag{6.82}$$

having entanglement rank bounded by r, Proposition 6.16 implies that

$$\langle \tau, \rho\rangle = \frac{1}{n}\big\langle \operatorname{vec}(\mathbb{1})\operatorname{vec}(\mathbb{1})^*, \rho\big\rangle \leq \frac{r}{n}. \tag{6.83}$$

One therefore has that every state of bounded entanglement rank must have a proportionately small inner product with the state τ.

6.1.2 Separable maps and the LOCC paradigm

Separable maps are defined in an analogous way to separable operators, reflecting the natural correspondence between completely positive maps and positive semidefinite operators. The resulting notion of separability for maps, including channels, is algebraic in nature; and it cannot be said that it is directly motivated from a physical or operational viewpoint.

This notion of separability for channels is, however, closely connected to the more operationally motivated notion of channels implementable by *local operations and classical communication* (or *LOCC* for short). An LOCC channel is a channel that can be implemented by two individuals whose local actions are unrestricted (corresponding to arbitrary channels or measurements), but whose communications with one another are restricted to be classical. This paradigm provides a foundation from which properties of entanglement are commonly studied, particularly in settings in which entanglement is viewed as a resource for information processing.

Separable map and channels

As suggested above, the notion of separability for maps is defined in an analogous way to separability for operators. The following definition states this in more precise terms.

Definition 6.18. Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be complex Euclidean spaces. The set

$$\text{SepCP}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W}) \tag{6.84}$$

is defined as the set of all completely positive maps of the form

$$\Xi \in \text{CP}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Z} \otimes \mathcal{W}) \tag{6.85}$$

for which there exists an alphabet Σ and collections of completely positive maps

$$\{\Phi_a : a \in \Sigma\} \subset \text{CP}(\mathcal{X}, \mathcal{Z}) \quad \text{and} \quad \{\Psi_a : a \in \Sigma\} \subset \text{CP}(\mathcal{Y}, \mathcal{W}) \tag{6.86}$$

such that

$$\Xi = \sum_{a \in \Sigma} \Phi_a \otimes \Psi_a. \tag{6.87}$$

Elements of the set $\text{SepCP}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$ are called *separable maps*.

As the following simple proposition states, separable maps are precisely those completely positive maps having Kraus representations in which the individual Kraus operators are tensor products of operators. A direct proof of this proposition is obtained by considering Kraus representations of the maps Φ_a and Ψ_a in Definition 6.18, along the same lines as the proof of Proposition 6.6.

Proposition 6.19. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be complex Euclidean spaces and let*

$$\Xi \in \mathrm{CP}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Z} \otimes \mathcal{W}) \tag{6.88}$$

be a completely positive map. It holds that $\Phi \in \mathrm{SepCP}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$ if and only if there exists an alphabet Σ and collections of operators

$$\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Z}) \quad \text{and} \quad \{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y}, \mathcal{W}) \tag{6.89}$$

such that

$$\Xi(X) = \sum_{a \in \Sigma} (A_a \otimes B_a) X (A_a \otimes B_a)^* \tag{6.90}$$

for every operator $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$.

Another straightforward proposition regarding separable maps is the following proposition, which implies that the set of all separable maps is closed under composition. Like the previous proposition, it may be verified directly.

Proposition 6.20. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $\mathcal{W}$, $\mathcal{U}$, and $\mathcal{V}$ be complex Euclidean spaces, and suppose that Φ and Ψ are separable maps of the form*

$$\Phi \in \mathrm{SepCP}(\mathcal{X}, \mathcal{U} : \mathcal{Y}, \mathcal{V}) \quad \text{and} \quad \Psi \in \mathrm{SepCP}(\mathcal{U}, \mathcal{Z} : \mathcal{V}, \mathcal{W}). \tag{6.91}$$

It holds that the composition $\Psi\Phi$ is separable:

$$\Psi\Phi \in \mathrm{SepCP}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W}). \tag{6.92}$$

Similar to the analogous case for states, one defines the set of separable channels by simply restricting the definition of separability for completely positive maps to channels.

Definition 6.21. For complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$, one defines

$$\begin{aligned} &\mathrm{SepC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W}) \\ &\quad = \mathrm{SepCP}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W}) \cap \mathrm{C}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Z} \otimes \mathcal{W}). \end{aligned} \tag{6.93}$$

Elements of the set $\mathrm{SepC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$ are referred to as *separable channels*.

It should be noted that, unlike the analogous case of states, separable channels need not be equal to convex combinations of product channels, as the following example illustrates.

Example 6.22. Let $\Sigma = \{0,1\}$ denote the binary alphabet, let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ all be equal to $\mathbb{C}^{\Sigma}$, and define a channel $\Xi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Z} \otimes \mathcal{W})$ by the equation

$$\Xi(E_{a,b} \otimes E_{c,d}) = \begin{cases} E_{a,a} \otimes E_{a,a} & \text{if } a = b \text{ and } c = d \\ 0 & \text{if } a \neq b \text{ or } c \neq d, \end{cases} \tag{6.94}$$

holding for all $a, b, c, d \in \Sigma$. It is the case that Ξ is a separable channel, meaning that $\Xi \in \mathrm{SepC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$. Indeed, one may write

$$\Xi = \Phi_0 \otimes \Psi_0 + \Phi_1 \otimes \Psi_1 \tag{6.95}$$

for completely positive maps defined as follows:

$$\begin{aligned} \Phi_0(X) &= \langle E_{0,0}, X \rangle E_{0,0}, \quad \Psi_0(X) = \mathrm{Tr}(X) E_{0,0}, \\ \Phi_1(X) &= \langle E_{1,1}, X \rangle E_{1,1}, \quad \Psi_1(X) = \mathrm{Tr}(X) E_{1,1}, \end{aligned} \tag{6.96}$$

for every $X \in \mathrm{L}\big(\mathbb{C}^{\Sigma}\big)$.

It is not possible, however, to express the channel Ξ in the form

$$\Xi = \sum_{a \in \Gamma} p(a) \Phi_a \otimes \Psi_a \tag{6.97}$$

for any choice of an alphabet Γ, a probability vector $p \in \mathcal{P}(\Gamma)$, and two collections of channels

$$\{\Phi_a : a \in \Gamma\} \subseteq \mathrm{C}(\mathcal{X}, \mathcal{Z}) \quad \text{and} \quad \{\Psi_a : a \in \Gamma\} \subseteq \mathrm{C}(\mathcal{Y}, \mathcal{W}). \tag{6.98}$$

To verify this claim, consider the fact that

$$\Xi(E_{0,0} \otimes \rho) = E_{0,0} \otimes E_{0,0} \quad \text{and} \quad \Xi(E_{1,1} \otimes \rho) = E_{1,1} \otimes E_{1,1} \tag{6.99}$$

for every density operator $\rho \in \mathrm{D}(\mathcal{Y})$. If it were the case that (6.97) were true for each Φ_a and Ψ_a being a channel, then one would necessarily have

$$\sum_{a \in \Gamma} p(a) \Phi_a(E_{0,0}) \otimes \Psi_a(\rho) = E_{0,0} \otimes E_{0,0}, \tag{6.100}$$

and therefore, by tracing over the space $\mathcal{Z}$,

$$\sum_{a\in\Sigma} p(a)\Psi_a(\rho) = E_{0,0} \tag{6.101}$$

for every $\rho \in \mathrm{D}(\mathcal{Y})$. By similar reasoning, it would simultaneously hold that

$$\sum_{a\in\Sigma} p(a)\Phi_a(E_{1,1}) \otimes \Psi_a(\rho) = E_{1,1} \otimes E_{1,1}, \tag{6.102}$$

and therefore

$$\sum_{a\in\Sigma} p(a)\Psi_a(\rho) = E_{1,1} \tag{6.103}$$

for every $\rho \in \mathrm{D}(\mathcal{Y})$. The equations (6.101) and (6.103) are in contradiction, implying that Ξ is not equal to a convex combination of product channels.

Intuitively speaking, the situation represented by the previous example is quite simple. Channels that can be expressed as a convex combination of product channels correspond to transformations that may be implemented by means of *local operations and shared randomness*—no communication is needed to implement them, and such channels do not allow for a direct causal relationship to hold among the input and output systems across the bipartition with respect to which separability is considered. The channel Ξ, on the other hand, induces a direct causal relationship of this form.

As the following proposition states, a given completely positive map is separable if and only if its Choi representation is separable, with respect to the natural bipartition of the tensor product space over which it is defined.

Proposition 6.23. *Let $\Xi \in \mathrm{CP}(\mathcal{X}\otimes\mathcal{Y}, \mathcal{Z}\otimes\mathcal{W})$ be a completely positive map, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$, and define an isometry*

$$V \in \mathrm{U}(\mathcal{Z}\otimes\mathcal{W}\otimes\mathcal{X}\otimes\mathcal{Y},\ \mathcal{Z}\otimes\mathcal{X}\otimes\mathcal{W}\otimes\mathcal{Y}) \tag{6.104}$$

by the equation

$$V\,\mathrm{vec}(A\otimes B) = \mathrm{vec}(A)\otimes\mathrm{vec}(B) \tag{6.105}$$

holding for all operators $A \in \mathrm{L}(\mathcal{X},\mathcal{Z})$ and $B \in \mathrm{L}(\mathcal{Y},\mathcal{W})$. It holds that

$$\Xi \in \mathrm{SepCP}(\mathcal{X},\mathcal{Z} : \mathcal{Y},\mathcal{W}) \tag{6.106}$$

if and only if

$$VJ(\Xi)V^* \in \mathrm{Sep}(\mathcal{Z}\otimes\mathcal{X} : \mathcal{W}\otimes\mathcal{Y}). \tag{6.107}$$

Proof. Assume first that Ξ is a separable map. By Proposition 6.19, there must exist an alphabet Σ and two collections of operators,

$$\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Z}) \quad \text{and} \quad \{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y}, \mathcal{W}), \tag{6.108}$$

such that

$$\Xi(X) = \sum_{a \in \Sigma} (A_a \otimes B_a) X (A_a \otimes B_a)^* \tag{6.109}$$

for every operator $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$. The Choi representation of Ξ is therefore given by

$$J(\Xi) = \sum_{a \in \Sigma} \operatorname{vec}(A_a \otimes B_a) \operatorname{vec}(A_a \otimes B_a)^*, \tag{6.110}$$

so that

$$VJ(\Xi)V^* = \sum_{a \in \Sigma} \operatorname{vec}(A_a) \operatorname{vec}(A_a)^* \otimes \operatorname{vec}(B_a) \operatorname{vec}(B_a)^*, \tag{6.111}$$

which is evidently contained in $\mathrm{Sep}(\mathcal{Z} \otimes \mathcal{X} : \mathcal{W} \otimes \mathcal{Y})$.

Conversely, if $VJ(\Xi)V^*$ is separable, then it must be possible to express this operator in the form (6.111) for some choice of an alphabet Σ and two collections of operators as in (6.108). It therefore follows that (6.110) is a Choi representation of Ξ, so that (6.109) holds for all $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$. The map Ξ is therefore separable, which completes the proof. □

Remark 6.24. The isometry V defined in Proposition 6.23 may alternatively be defined by the action

$$V(z \otimes w \otimes x \otimes y) = z \otimes x \otimes w \otimes y, \tag{6.112}$$

for every choice of vectors $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, and $w \in \mathcal{W}$. In words, this isometry represents a permutation of tensor factors, allowing a relationship concerning separability with respect to a particular bipartition to be stated precisely.

It is not uncommon in the theory of quantum information literature that statements of this nature are made without an explicit mention of such an isometry. This can sometimes simplify expressions and generally does not lead to any confusion—the isometry can usually be taken as being implicit, particularly in cases when the underlying complex Euclidean spaces have distinct names. In the interest of clarity and formality, however, this book will always represent such permutations of tensor factors explicitly.

Separable channels are not capable of creating entanglement; when a separable channel is applied to a separable state, the output is necessarily another separable state. More generally, separable maps cannot cause an increase entanglement rank, as the following theorem establishes.

Theorem 6.25. *Let $\Xi \in \mathrm{SepCP}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$ be a separable map, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$. For every positive integer r and every operator $P \in \mathrm{Ent}_r(\mathcal{X} : \mathcal{Y})$, it holds that $\Xi(P) \in \mathrm{Ent}_r(\mathcal{Z} : \mathcal{W})$.*

Proof. For an operator $P \in \mathrm{Ent}_r(\mathcal{X} : \mathcal{Y})$ having entanglement rank bounded by r, there must exist an alphabet Γ and a collection of operators

$$\{X_b \,:\, b \in \Gamma\} \subset \mathrm{L}(\mathcal{Y}, \mathcal{X}), \tag{6.113}$$

satisfying $\mathrm{rank}(X_b) \leq r$ for every $b \in \Gamma$, such that

$$P = \sum_{b \in \Gamma} \mathrm{vec}(X_b)\,\mathrm{vec}(X_b)^*. \tag{6.114}$$

Proposition 6.19, it follows that

$$\begin{aligned} \Xi(P) &= \sum_{a \in \Sigma} \sum_{b \in \Gamma} (A_a \otimes B_a)\,\mathrm{vec}(X_b)\,\mathrm{vec}(X_b)^* (A_a \otimes B_a)^* \\ &= \sum_{a \in \Sigma} \sum_{b \in \Gamma} \mathrm{vec}\big(A_a X_b B_a^{\mathsf{T}}\big)\,\mathrm{vec}\big(A_a X_b B_a^{\mathsf{T}}\big)^* \end{aligned} \tag{6.115}$$

for some choice of an alphabet Σ and two collections of operators

$$\{A_a \,:\, a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}, \mathcal{Z}) \quad \text{and} \quad \{B_a \,:\, a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y}, \mathcal{W}). \tag{6.116}$$

For every $a \in \Sigma$ and $b \in \Gamma$, it holds that

$$\mathrm{rank}\big(A_a X_b B_a^{\mathsf{T}}\big) \leq \mathrm{rank}(X_b) \leq r, \tag{6.117}$$

and therefore $\Xi(P) \in \mathrm{Ent}_r(\mathcal{Z} : \mathcal{W})$, as required. □

Corollary 6.26. *Let $\Xi \in \mathrm{SepCP}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$ be a separable map, for complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$. For every separable operator $P \in \mathrm{Sep}(\mathcal{X} : \mathcal{Y})$, it holds that $\Xi(P)$ is also separable: $\Xi(P) \in \mathrm{Sep}(\mathcal{Z} : \mathcal{W})$.*

LOCC channels

As was stated at the beginning of the present subsection, LOCC channels represent transformations of quantum states that may be implemented by two individuals that communicate with one another classically and perform quantum channels and measurements on registers they hold locally.

For instance, one individual may apply a combination of channels and measurements to a collection of registers in their possession and transmit the measurement outcomes to the other individual. Upon receiving this transmission, the other individual may apply channels and measurements depending on the communicated measurement outcomes to registers in their possession. In general, LOCC channels may represent the cumulative effect of composing any finite[2] number of transformations of this sort.

The following definition formalizes this notion. Naturally, it is possible to generalize this definition to three or more individuals, although this will not be done in this book.

Definition 6.27. Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be complex Euclidean spaces and let

$$\Xi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Z} \otimes \mathcal{W}) \tag{6.118}$$

be a channel. The channel Ξ is an *LOCC channel* under these conditions:

1. The channel Ξ is a *one-way right LOCC channel* if and only if there exists an alphabet Σ and a collection

   $$\{\Phi_a : a \in \Sigma\} \subset \mathrm{CP}(\mathcal{X}, \mathcal{Z}) \tag{6.119}$$

   of completely positive maps satisfying

   $$\sum_{a \in \Sigma} \Phi_a \in \mathrm{C}(\mathcal{X}, \mathcal{Z}), \tag{6.120}$$

   along with a collection

   $$\{\Psi_a : a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{Y}, \mathcal{W}) \tag{6.121}$$

   of channels, such that

   $$\Xi = \sum_{a \in \Sigma} \Phi_a \otimes \Psi_a. \tag{6.122}$$

[2] One may consider variants of the definition that allow for an unbounded number of classical transmissions that terminate with probability 1 according to a chosen stopping rule. Only the finite case is considered in this book for simplicity.

2. The channel Ξ is a *one-way left LOCC channel* if and only if there exists an alphabet Σ and a collection

$$\{\Psi_a : a \in \Sigma\} \subset \mathrm{CP}(\mathcal{Y}, \mathcal{W}) \tag{6.123}$$

of completely positive maps satisfying

$$\sum_{a \in \Sigma} \Psi_a \in \mathrm{C}(\mathcal{Y}, \mathcal{W}), \tag{6.124}$$

along with a collection

$$\{\Phi_a : a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{X}, \mathcal{Z}) \tag{6.125}$$

of channels, such that (6.122) holds.

3. The channel Ξ is an *LOCC channel* if and only if it is equal to a finite composition of one-way left and one-way right LOCC channels. That is, either Ξ is a one-way left LOCC channel, a one-way right LOCC channel, or there exists an integer $m \geq 2$, complex Euclidean spaces $\mathcal{U}_1, \ldots, \mathcal{U}_{m-1}$ and $\mathcal{V}_1, \ldots, \mathcal{V}_{m-1}$, and channels

$$\begin{aligned} \Xi_1 &\in \mathrm{C}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{U}_1 \otimes \mathcal{V}_1), \\ \Xi_2 &\in \mathrm{C}(\mathcal{U}_1 \otimes \mathcal{V}_1, \mathcal{U}_2 \otimes \mathcal{V}_2), \\ &\vdots \\ \Xi_m &\in \mathrm{C}(\mathcal{U}_{m-1} \otimes \mathcal{V}_{m-1}, \mathcal{Z} \otimes \mathcal{W}), \end{aligned} \tag{6.126}$$

each of which is either a one-way left LOCC channel or a one-way right LOCC channel, such that Ξ is equal to the composition $\Xi = \Xi_m \cdots \Xi_1$.

The collection of all such LOCC channels is denoted $\mathrm{LOCC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$.

Remark 6.28. In the definition above, one-way left and one-way right LOCC channels represent channels that can be implemented by local operations and one-way classical communication. In both cases, the channel Ξ may be viewed as having resulted from actions performed by two individuals, Alice and Bob. Alice begins with a register X and Bob begins with Y, and as a result of their actions these registers are transformed into Z and W, respectively.

In the case of a one-way right LOCC channel Ξ, the communication is from Alice to Bob (moving to the *right*, assuming Alice is on the left and

Bob is on the right), with the alphabet Σ representing the set of possible classical messages that may be transmitted. Alice's actions are described by a collection of completely positive maps

$$\{\Phi_a : a \in \Sigma\} \subset \mathrm{CP}(\mathcal{X}, \mathcal{Z}) \tag{6.127}$$

that satisfies the constraint

$$\sum_{a \in \Sigma} \Phi_a \in \mathrm{C}(\mathcal{X}, \mathcal{Z}). \tag{6.128}$$

In essence, this collection specifies a quantum instrument (q.v. Section 2.3.2). Assuming the classical communication is represented by a classical register V having associated complex Euclidean space $\mathcal{V} = \mathbb{C}^{\Sigma}$, Alice's action would be described by the channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Z} \otimes \mathcal{V})$ defined by

$$\Phi(X) = \sum_{a \in \Sigma} \Phi_a(X) \otimes E_{a,a} \tag{6.129}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The register V is sent to Bob, who observes its classical state (or, equivalently, measures V with respect to the standard basis) and transforms his register Y into W according to the channel $\Psi_a \in \mathrm{C}(\mathcal{Y}, \mathcal{W})$, for $a \in \Sigma$ being the classical state of V that was observed. Assuming that the register V is discarded after Bob applies the appropriate channel, the combined actions of Alice and Bob are described by Ξ.

For a one-way left LOCC channel Ξ, the situation is similar, with the roles of Alice and Bob switched.

It is apparent from Definition 6.27, together with the fact that separable channels are closed under composition (Proposition 6.20), that every LOCC channel is a separable channel.

Proposition 6.29. *For every choice of complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$, it holds that*

$$\mathrm{LOCC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W}) \subseteq \mathrm{SepC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W}). \tag{6.130}$$

6.1.3 Separable and LOCC measurements

As was explained in Section 2.3.1, one may associate a quantum-to-classical channel with each measurement, with the classical output of the channel representing the outcome of the measurement. Through an identification of this sort, the notions of separable and LOCC channels may be extended to measurements.

Definitions of separable and LOCC measurements

The following definition of separable and LOCC measurements refers to an association of quantum-to-classical channels with measurements that has been adapted to a bipartite setting.

Definition 6.30. Let Σ be an alphabet, let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and let $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ be a measurement. Define complex Euclidean spaces $\mathcal{Z} = \mathbb{C}^{\Sigma}$ and $\mathcal{W} = \mathbb{C}^{\Sigma}$, and define a channel

$$\Phi_{\mu} \in \mathrm{C}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Z} \otimes \mathcal{W}) \tag{6.131}$$

as

$$\Phi_{\mu}(X) = \sum_{a \in \Sigma} \langle \mu(a), X \rangle \, E_{a,a} \otimes E_{a,a} \tag{6.132}$$

for every $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$. The measurement μ is a *separable measurement* if and only if

$$\Phi_{\mu} \in \mathrm{SepC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W}), \tag{6.133}$$

and μ is an *LOCC measurement* if and only if

$$\Phi_{\mu} \in \mathrm{LOCC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W}). \tag{6.134}$$

For a given measurement μ, the channel Φ_{μ} specified in Definition 6.30 is similar to the quantum-to-classical channel one would normally associate with μ, except that two copies of the measurement outcome are produced rather than one. In a bipartite setting, this is natural way of associating a quantum-to-classical channel with a measurement. If this measurement is performed on a pair of registers (X, Y) by two individuals, Alice and Bob, where it is assumed that Alice holds X and Bob holds Y, the channel Φ_{μ} represents the measurement μ under the assumption that both individuals learn the measurement outcome after the measurement is performed.

One alternative to Definition 6.30 is to replace the channel Φ_{μ} by the quantum-to-classical channel that would ordinarily be associated with the measurement μ, along with a specification of which side of the bipartition the measurement outcome is to fall (requiring this channel to be separable or LOCC, as in the stated definition). In essence, with respect to a situation in which Alice and Bob are performing the measurement μ as suggested above, such a definition specifies which of the two individuals obtains the

measurement outcome. This alternative creates an artificial asymmetry in the definition, but is equivalent to Definition 6.30.

With respect to Definition 6.30, the separability of a given measurement is equivalent to the constraint that each measurement operator is separable, as the following proposition states.

Proposition 6.31. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let Σ be an alphabet, and let μ be a measurement of the form $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$. It holds that μ is a separable measurement if and only if $\mu(a) \in \mathrm{Sep}(\mathcal{X} : \mathcal{Y})$ for every $a \in \Sigma$.*

Proof. Consider the Choi representation of the mapping Φ_μ, as specified in Definition 6.30, which is given by

$$J(\Phi_\mu) = \sum_{a \in \Sigma} E_{a,a} \otimes E_{a,a} \otimes \overline{\mu(a)}. \tag{6.135}$$

Along similar lines to the statement of Proposition 6.23, let

$$V \in \mathrm{U}(\mathcal{Z} \otimes \mathcal{W} \otimes \mathcal{X} \otimes \mathcal{Y},\, \mathcal{Z} \otimes \mathcal{X} \otimes \mathcal{W} \otimes \mathcal{Y}) \tag{6.136}$$

be the isometry defined by the equation

$$V \,\mathrm{vec}(A \otimes B) = \mathrm{vec}(A) \otimes \mathrm{vec}(B) \tag{6.137}$$

holding for all operators $A \in \mathrm{L}(\mathcal{X}, \mathcal{Z})$ and $B \in \mathrm{L}(\mathcal{Y}, \mathcal{W})$. If it is the case that $\mu(a) \in \mathrm{Sep}(\mathcal{X} \otimes \mathcal{Y})$ for every $a \in \Sigma$, then it follows directly that

$$V J(\Phi_\mu) V^* \in \mathrm{Sep}(\mathcal{Z} \otimes \mathcal{X} : \mathcal{W} \otimes \mathcal{Y}), \tag{6.138}$$

which implies that μ is a separable measurement by Proposition 6.23.

Now suppose that μ is a separable measurement, so that (6.138) holds. Define a mapping $\Xi_a \in \mathrm{T}(\mathcal{Z} \otimes \mathcal{X} \otimes \mathcal{W} \otimes \mathcal{Y}, \mathcal{X} \otimes \mathcal{Y})$, for each $a \in \Sigma$, as

$$\Xi_a(X) = \big((e_a^* \otimes \mathbb{1}_\mathcal{X}) \otimes (e_a^* \otimes \mathbb{1}_\mathcal{Y})\big) X \big((e_a \otimes \mathbb{1}_\mathcal{X}) \otimes (e_a \otimes \mathbb{1}_\mathcal{Y})\big) \tag{6.139}$$

for all $X \in \mathrm{L}(\mathcal{Z} \otimes \mathcal{X} \otimes \mathcal{W} \otimes \mathcal{Y})$. It is evident from this definition that Ξ_a is a separable mapping for each $a \in \Sigma$, meaning

$$\Xi_a \in \mathrm{SepCP}(\mathcal{Z} \otimes \mathcal{X}, \mathcal{X} : \mathcal{W} \otimes \mathcal{Y}, \mathcal{Y}). \tag{6.140}$$

It holds that

$$\overline{\mu(a)} = \Xi_a\big(V J(\Phi_\mu) V^*\big) \tag{6.141}$$

for each $a \in \Sigma$, from which it follows that

$$\overline{\mu(a)} \in \text{Sep}(\mathcal{X} : \mathcal{Y}) \tag{6.142}$$

by Corollary 6.26. This is equivalent to $\mu(a) \in \text{Sep}(\mathcal{X} : \mathcal{Y})$ for each $a \in \Sigma$, as the entry-wise complex conjugate of every separable operator is evidently separable, which completes the proof. □

For two complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, along with an alphabet Σ, it is the case that the set of all separable measurements of the form

$$\mu : \Sigma \to \text{Pos}(\mathcal{X} \otimes \mathcal{Y}) \tag{6.143}$$

is a proper subset of the set of all measurements of the same form (aside from the trivial cases in which one of $\dim(\mathcal{X})$, $\dim(\mathcal{Y})$, or $|\Sigma|$ equals 1). As every LOCC channel is separable, it follows that every LOCC measurement is a separable measurement.

One-way LOCC measurements

An interesting restricted type of LOCC measurement is one in which only *one-way communication* is permitted. The following definition formalizes this type of measurement.

Definition 6.32. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let Σ be an alphabet, and let

$$\mu : \Sigma \to \text{Pos}(\mathcal{X} \otimes \mathcal{Y}) \tag{6.144}$$

be a measurement. The measurement μ is a *one-way LOCC measurement* if and only if either of the following two conditions is met:

1. There exists an alphabet Γ and a measurement $\nu : \Gamma \to \text{Pos}(\mathcal{X})$, along with a measurement $\pi_b : \Sigma \to \text{Pos}(\mathcal{Y})$ for each $b \in \Gamma$, such that the equation

   $$\mu(a) = \sum_{b \in \Gamma} \nu(b) \otimes \pi_b(a) \tag{6.145}$$

   holds for every $a \in \Sigma$. In this case the measurement μ is said to be an *one-way right LOCC measurement*.

2. There exists an alphabet Γ and a measurement $\nu : \Gamma \to \mathrm{Pos}(\mathcal{Y})$, along with a measurement $\pi_b : \Sigma \to \mathrm{Pos}(\mathcal{X})$ for each $b \in \Gamma$, such that the equation
$$\mu(a) = \sum_{b \in \Gamma} \pi_b(a) \otimes \nu(b) \tag{6.146}$$
holds for every $a \in \Sigma$. In this case the measurement μ is said to be a *one-way left LOCC measurement*.

Limitations on state discrimination by separable measurements

One may consider the problem of state discrimination, as was discussed in Chapter 3, in which measurements are restricted to be separable or LOCC measurements. Many examples of sets of orthogonal pure states that cannot be distinguished without error by separable or LOCC measurements are known. The following theorem provides one class of examples, and implies that there exist relatively small sets of orthogonal pure states having this characteristic.

Theorem 6.33. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces of equal dimension n, let*
$$\{U_1, \ldots, U_m\} \in \mathrm{U}(\mathcal{Y}, \mathcal{X}) \tag{6.147}$$
be a set of pairwise orthogonal isometries, meaning that $\langle U_j, U_k \rangle = 0$ for $j \neq k$, and let $u_k \in \mathcal{X} \otimes \mathcal{Y}$ be the vector defined as
$$u_k = \frac{1}{\sqrt{n}} \mathrm{vec}(U_k) \tag{6.148}$$
for each $k \in \{1, \ldots, m\}$. For every separable measurement of the form
$$\mu : \{1, \ldots, m\} \to \mathrm{Sep}(\mathcal{X} : \mathcal{Y}) \tag{6.149}$$
it holds that
$$\sum_{k=1}^{m} \langle \mu(k), u_k u_k^* \rangle \leq n. \tag{6.150}$$

Proof. Under the assumption that μ is a separable measurement, one may write
$$\mu(k) = \sum_{a \in \Sigma} P_{k,a} \otimes Q_{k,a} \tag{6.151}$$

for each $k \in \{1, \ldots, m\}$, for some choice of an alphabet Σ and being some alphabet and collections

$$\begin{aligned} &\{P_{k,a} \,:\, k \in \{1, \ldots, m\},\, a \in \Sigma\} \subset \mathrm{Pos}(\mathcal{X}) \\ &\{Q_{k,a} \,:\, k \in \{1, \ldots, m\},\, a \in \Sigma\} \subset \mathrm{Pos}(\mathcal{Y}) \end{aligned} \tag{6.152}$$

of positive semidefinite operators. (There is no generality lost in using the same alphabet Σ in the expression (6.151) for each choice of k, as one is free to choose Σ to be as large as is needed, and to set $P_{k,a} = 0$ or $Q_{k,a}$ for some choices of k and a as necessary.) It holds that

$$\begin{aligned} \langle \mu(k), \mathrm{vec}(U_k)\,\mathrm{vec}(U_k)^* \rangle &= \sum_{a \in \Sigma} \langle U_k, P_{k,a} U_k Q_{k,a}^* \rangle \\ &\leq \sum_{a \in \Sigma} \| P_{k,a} U_k Q_{k,a}^* \|_1 \leq \sum_{a \in \Sigma} \| P_{k,a} \|_1 \| U_k Q_{k,a}^* \|_1 \\ &= \sum_{a \in \Sigma} \mathrm{Tr}(P_{k,a})\,\mathrm{Tr}(Q_{k,a}) = \mathrm{Tr}(\mu(k)), \end{aligned} \tag{6.153}$$

and therefore

$$\sum_{k=1}^{m} \langle \mu(k), \mathrm{vec}(U_k)\,\mathrm{vec}(U_k)^* \rangle \leq \sum_{k=1}^{m} \mathrm{Tr}(\mu(k)) = n^2. \tag{6.154}$$

The theorem follows by dividing both sides of this inequality by n. □

For any set of pure states $\{u_1, \ldots, u_m\}$ as described by this theorem, for which $m > n$, one therefore has that

$$\frac{1}{m} \sum_{k=1}^{m} \langle \mu(k), u_k u_k^* \rangle \leq \frac{n}{m} < 1. \tag{6.155}$$

Consequently, for one of these m states being uniformly selected at random, any separable measurement that aims to discriminate these states must err with probability strictly greater than 0.

LOCC discrimination of any pair of orthogonal pure states

Although Theorem 6.33 establishes that there exist relatively small sets of orthonormal pure states that cannot be perfectly discriminated by separable measurements, the same cannot be said about *pairs* of orthonormal pure states. Indeed, every pair of orthonormal pure states can be discriminated without error by a one-way LOCC measurement. The following lemma is used to prove this fact.

Lemma 6.34. *Let $\mathcal{X}$ be a complex Euclidean space, let $X \in \mathrm{L}(\mathcal{X})$ be an operator satisfying $\mathrm{Tr}(X) = 0$, and let $n = \dim(\mathcal{X})$. There exists an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$ such that $x_k^* X x_k = 0$ for all $k \in \{1, \ldots, n\}$.*

Proof. The proof is by induction on n. The base case $n = 1$ is immediate, so it will be assumed that $n \geq 2$ for the rest of the proof. It will also be assumed that $\mathcal{X} = \mathbb{C}^n$, which causes no loss of generality.

For every integer $k \in \{1, \ldots, n\}$, it holds that $\lambda_k(X) \in \mathcal{N}(X)$, where $\mathcal{N}(X)$ denotes the numerical range of X. By the Toeplitz–Hausdorff theorem (Theorem 3.56), the numerical range is convex, and therefore

$$0 = \frac{1}{n}\mathrm{Tr}(X) = \frac{1}{n}\sum_{k=1}^{n} \lambda_k(X) \in \mathcal{N}(X). \tag{6.156}$$

By the definition of the numerical range, there must therefore exist a unit vector $x_n \in \mathcal{X}$ such that $x_n^* X x_n = 0$.

Let $V \in \mathrm{U}(\mathbb{C}^{n-1}, \mathbb{C}^n)$ be any isometry that satisfies $x_n \perp \mathrm{im}(V)$, which is equivalent to $VV^* = \mathbb{1} - x_n x_n^*$. It holds that

$$\mathrm{Tr}(V^* X V) = \mathrm{Tr}((\mathbb{1} - x_n x_n^*)X) = \mathrm{Tr}(X) - x_n^* X x_n = 0. \tag{6.157}$$

As $V^* X V \in \mathrm{L}(\mathbb{C}^{n-1})$, the hypothesis of induction implies that there exist an orthonormal basis $\{u_1, \ldots, u_{n-1}\}$ of $\mathbb{C}^{n-1}$ such that

$$u_k^*(V^* X V)u_k = 0 \tag{6.158}$$

for all $k \in \{1, \ldots, n-1\}$. Define $x_k = Vu_k$ for each $k \in \{1, \ldots, n-1\}$, and observe that $\{x_1, \ldots, x_{n-1}\}$ is an orthonormal set, with each element x_k of this set satisfying $x_k^* X x_k = 0$. As V is an isometry and $x_n \perp \mathrm{im}(X)$, it follows that $\{x_1, \ldots, x_n\}$ is an orthonormal basis of $\mathcal{X}$ having the property stated by the lemma. □

Theorem 6.35. *Let $u_0, u_1 \in \mathcal{X} \otimes \mathcal{Y}$ be orthogonal unit vectors, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. There exists a one-way LOCC measurement*

$$\mu : \{0, 1\} \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}) \tag{6.159}$$

such that

$$\langle \mu(0), u_0 u_0^* \rangle = 1 = \langle \mu(1), u_1 u_1^* \rangle. \tag{6.160}$$

Proof. Let $n = \dim(\mathcal{Y})$ and let $A_0, A_1 \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be the unique operators satisfying $u_0 = \mathrm{vec}(A_0)$ and $u_1 = \mathrm{vec}(A_1)$. The orthogonality of the vectors u_0 and u_1 is equivalent to the condition $\mathrm{Tr}(A_0^* A_1) = 0$. By Lemma 6.34, there exists an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{Y}$ with the property that $x_k^* A_0^* A_1 x_k = 0$, which is equivalent to the condition that

$$\left\langle A_0 x_k x_k^* A_0^*, A_1 x_k x_k^* A_1^* \right\rangle = 0, \tag{6.161}$$

for every $k \in \{1, \ldots, n\}$.

Define a measurement $\nu : \{1, \ldots, n\} \to \mathrm{Pos}(\mathcal{Y})$ as

$$\nu(k) = \overline{x_k} x_k^{\mathsf{T}} \tag{6.162}$$

for each $k \in \{1, \ldots, n\}$. By the equation (6.161), one has that there must exist a measurement $\pi_k : \{0, 1\} \to \mathrm{Pos}(\mathcal{X})$, for each $k \in \{1, \ldots, n\}$, such that

$$\left\langle \pi_k(0), A_1 x_k x_k^* A_1^* \right\rangle = 0 = \left\langle \pi_k(1), A_0 x_k x_k^* A_0^* \right\rangle. \tag{6.163}$$

Finally, define $\mu : \{0, 1\} \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ as

$$\mu(a) = \sum_{k=1}^{n} \pi_k(a) \otimes \nu(k) \tag{6.164}$$

for each $a \in \{0, 1\}$, which is a one-way measurement with respect to the second condition of Definition 6.32. It holds that

$$\begin{aligned} \left\langle \mu(0), u_1 u_1^* \right\rangle &= \sum_{k=1}^{n} \left\langle \pi_k(0), (\mathbb{1} \otimes x_k^{\mathsf{T}}) \,\mathrm{vec}(A_1)\,\mathrm{vec}(A_1)^* (\mathbb{1} \otimes \overline{x_k}) \right\rangle \\ &= \sum_{k=1}^{n} \left\langle \pi_k(0), A_1 x_k x_k^* A_1^* \right\rangle = 0, \end{aligned} \tag{6.165}$$

and through a similar calculation one finds that $\left\langle \mu(1), u_0 u_0^* \right\rangle = 0$, which completes the proof. □

Remark 6.36. The preceding proof may be adapted in a straightforward way to prove that there exists a one-way LOCC measurement respecting the first condition of Definition 6.32, as opposed to the second, that satisfies the requirements of the theorem.

6.2 Manipulation of entanglement

As presented in the previous section, entanglement is defined as a lack of separability—for two complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, a bipartite state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ that is not contained in the set $\mathrm{SepD}(\mathcal{X} : \mathcal{Y})$ is entangled with respect to the bipartition between $\mathcal{X}$ and $\mathcal{Y}$. This definition is qualitative, in the sense that it does not provide a measure of how much entanglement is present in a given state or suggest how two entangled states might or might not relate to one another. The present section discusses such notions, and develops basic concepts and techniques relating to quantitative aspects of entanglement.

6.2.1 Entanglement transformation

The next theorem establishes a necessary and sufficient condition under which two individuals may transform one pure state into another by means of local operations and classical communication. The condition concerns the reductions of the initial and final pure states to one of the two individuals, requiring that the reduction of the initial state is majorized by the reduction of the final state. This condition is not only equivalent to the existence of an LOCC (or even a separable) channel transforming the initial state to the final state, but also implies that the transformation can be accomplished with one-way classical communication, from either of the two individuals to the other. The theorem offers a tool through which two fundamental ways of quantifying how much entanglement exists in a given state, called the *entanglement cost* and the *distillable entanglement*, may be analyzed for pure states.

Theorem 6.37 (Nielsen's theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $u, v \in \mathcal{X} \otimes \mathcal{Y}$ be unit vectors. The following statements are equivalent:*

1. $\mathrm{Tr}_{\mathcal{Y}}(uu^*) \prec \mathrm{Tr}_{\mathcal{Y}}(vv^*)$.
2. *There exists an alphabet Σ, a collection of operators $\{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y})$ satisfying*
$$\sum_{a \in \Sigma} B_a^* B_a = \mathbb{1}_{\mathcal{Y}}, \tag{6.166}$$
and a collection of unitary operators $\{U_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{X})$ such that
$$vv^* = \sum_{a \in \Sigma} (U_a \otimes B_a) uu^* (U_a \otimes B_a)^*. \tag{6.167}$$

3. *There exists an alphabet* Σ*, a collection of operators* $\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$ *satisfying*

$$\sum_{a\in\Sigma} A_a^* A_a = \mathbb{1}_{\mathcal{X}}, \tag{6.168}$$

and a collection of unitary operators $\{V_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{Y})$ *such that*

$$vv^* = \sum_{a\in\Sigma} (A_a \otimes V_a) uu^* (A_a \otimes V_a)^*. \tag{6.169}$$

4. *There exists a separable channel*[3] $\Phi \in \mathrm{SepC}(\mathcal{X} : \mathcal{Y})$ *such that* $vv^* = \Phi(uu^*)$.

Proof. Let $X, Y \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be the unique operators for which $u = \mathrm{vec}(X)$ and $v = \mathrm{vec}(Y)$, and let

$$X = \sum_{k=1}^{r} s_k x_k y_k^* \tag{6.170}$$

be a singular value decomposition of X, for $r = \mathrm{rank}(X)$.

Assume first that statement 1 holds, which is equivalent to $XX^* \prec YY^*$. There must therefore exist an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and a collection of unitary operators $\{W_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{X})$ such that

$$XX^* = \sum_{a\in\Sigma} p(a) W_a YY^* W_a^*. \tag{6.171}$$

Let $\mathcal{Z} = \mathbb{C}^{\Sigma}$ and define an operator $Z \in \mathrm{L}(\mathcal{Y} \otimes \mathcal{Z}, \mathcal{X})$ as

$$Z = \sum_{a\in\Sigma} \sqrt{p(a)}\, W_a Y \otimes e_a^*. \tag{6.172}$$

It holds that

$$ZZ^* = \sum_{a\in\Sigma} p(a) W_a YY^* W_a^* = XX^*, \tag{6.173}$$

and therefore Z and X agree on their singular values, and on the possible choices for their left singular vectors. It follows that one may write

$$Z = \sum_{k=1}^{r} s_k x_k w_k^* \tag{6.174}$$

for $\{w_1, \ldots, w_r\} \subset \mathcal{Y} \otimes \mathcal{Z}$ being an orthonormal collection of vectors. Let $V \in \mathrm{U}(\mathcal{Y}, \mathcal{Y} \otimes \mathcal{Z})$ be an isometry for which $Vy_k = w_k$ for all $k \in \{1, \ldots, r\}$, so that $XV^* = Z$.

[3] As one may expect, the notation $\mathrm{SepC}(\mathcal{X} : \mathcal{Y})$ is a shorthand for $\mathrm{SepC}(\mathcal{X}, \mathcal{X} : \mathcal{Y}, \mathcal{Y})$.

Now, define operators

$$U_a = W_a^* \quad \text{and} \quad B_a = (\mathbb{1}_{\mathcal{Y}} \otimes e_a^*)\overline{V} \tag{6.175}$$

for each $a \in \Sigma$. As V is an isometry, so too is $\overline{V}$, and therefore

$$\sum_{a\in\Sigma} B_a^* B_a = \sum_{a\in\Sigma} V^{\mathsf{T}}(\mathbb{1}_{\mathcal{Y}} \otimes E_{a,a})\overline{V} = V^{\mathsf{T}}\overline{V} = \mathbb{1}_{\mathcal{Y}}. \tag{6.176}$$

It holds that

$$W_a^* X B_a^{\mathsf{T}} = W_a^* X V^*(\mathbb{1}_{\mathcal{Y}} \otimes e_a) = W_a^* Z(\mathbb{1}_{\mathcal{Y}} \otimes e_a) = \sqrt{p(a)}\,Y \tag{6.177}$$

for each $a \in \Sigma$, and therefore

$$\begin{aligned}
&\sum_{a\in\Sigma} (U_a \otimes B_a) uu^* (U_a \otimes B_a)^* \\
&\qquad = \sum_{a\in\Sigma} \operatorname{vec}\bigl(W_a^* X B_a^{\mathsf{T}}\bigr) \operatorname{vec}\bigl(W_a^* X B_a^{\mathsf{T}}\bigr)^* \\
&\qquad = \sum_{a\in\Sigma} p(a) \operatorname{vec}(Y) \operatorname{vec}(Y)^* \\
&\qquad = vv^*.
\end{aligned} \tag{6.178}$$

It has been established that statement 1 implies statement 2.

The fact that statement 1 implies statement 3 is established by a similar argument with the roles of $\mathcal{X}$ and $\mathcal{Y}$ exchanged, along with the observation that $\operatorname{Tr}_{\mathcal{Y}}(uu^*) \prec \operatorname{Tr}_{\mathcal{Y}}(vv^*)$ is equivalent to $\operatorname{Tr}_{\mathcal{X}}(uu^*) \prec \operatorname{Tr}_{\mathcal{X}}(vv^*)$.

Statements 2 and 3 each imply statement 4 directly, as the mappings defined by the actions

$$\begin{aligned}
uu^* &\mapsto \sum_{a\in\Sigma} (U_a \otimes B_a) uu^* (U_a \otimes B_a)^*, \\
uu^* &\mapsto \sum_{a\in\Sigma} (A_a \otimes V_a) uu^* (A_a \otimes V_a)^*
\end{aligned} \tag{6.179}$$

are both separable channels.

Finally, assume statement 4 holds, letting $\Phi \in \operatorname{SepC}(\mathcal{X} : \mathcal{Y})$ be a fixed separable channel for which $\Phi(uu^*) = vv^*$. It will be proved that

$$\lambda(XX^*) \prec \lambda(YY^*); \tag{6.180}$$

by Theorem 4.33, this relation is equivalent to $XX^* \prec YY^*$, which in turn is equivalent to statement 1. Let $n = \dim(\mathcal{X})$, and observe that

$$\sum_{k=1}^{n} \lambda_k(XX^*) = \mathrm{Tr}(XX^*) = 1 = \mathrm{Tr}(YY^*) = \sum_{k=1}^{n} \lambda_k(YY^*), \tag{6.181}$$

by the assumption that u and v are unit vectors. By Theorem 4.30, one finds that the relation (6.180) will therefore follow from the inequality

$$\sum_{k=m}^{n} \lambda_k(YY^*) \leq \sum_{k=m}^{n} \lambda_k(XX^*) \tag{6.182}$$

holding for every choice of $m \in \{1, \ldots, n\}$.

By the separability of the channel Φ, there must exist an alphabet Σ and two collections of operators

$$\{A_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{X}) \quad \text{and} \quad \{B_a : a \in \Sigma\} \subset \mathrm{L}(\mathcal{Y}), \tag{6.183}$$

with $\{A_a \otimes B_a : a \in \Sigma\}$ being a set of Kraus operators of Φ, for which

$$vv^* = \sum_{a \in \Sigma} (A_a \otimes B_a) uu^* (A_a \otimes B_a)^*. \tag{6.184}$$

As vv^* is a rank-one operator, it follows that there must exist a probability vector $p \in \mathcal{P}(\Sigma)$ such that

$$(A_a \otimes B_a) uu^* (A_a \otimes B_a)^* = p(a) vv^*, \tag{6.185}$$

which is equivalent to

$$\mathrm{vec}\big(A_a X B_a^{\mathsf{T}}\big)\, \mathrm{vec}\big(A_a X B_a^{\mathsf{T}}\big)^* = p(a)\, \mathrm{vec}(Y)\, \mathrm{vec}(Y)^*, \tag{6.186}$$

for each $a \in \Sigma$. By taking the partial trace over $\mathcal{Y}$, it follows that

$$A_a X B_a^{\mathsf{T}} \overline{B_a} X^* A_a^* = p(a) YY^* \tag{6.187}$$

for each $a \in \Sigma$, and therefore

$$\sum_{k=m}^{n} \lambda_k\big(YY^*\big) = \sum_{k=m}^{n} \sum_{a \in \Sigma} \lambda_k\Big(A_a X B_a^{\mathsf{T}} \overline{B_a} X^* A_a^*\Big) \tag{6.188}$$

for each $m \in \{1, \ldots, n\}$.

Next, for each choice of $a \in \Sigma$ and $m \in \{1, \ldots, n\}$, let $\Pi_{a,m} \in \mathrm{Proj}(\mathcal{X})$ be the projection operator onto the orthogonal complement of the subspace of $\mathcal{X}$ spanned by the set $\{A_a x_1, \ldots, A_a x_{m-1}\}$, where one is to assume $x_k = 0$ for $k > r$. By the definition of these projection operators, it is evident that

$$\left\langle \Pi_{m,a}, A_a X B_a^{\mathsf{T}} \overline{B_a} X^* A_a^* \right\rangle = \left\langle \Pi_{m,a}, A_a X_m B_a^{\mathsf{T}} \overline{B_a} X_m^* A_a^* \right\rangle \tag{6.189}$$

for every $a \in \Sigma$ and $m \in \{1, \ldots, n\}$, where

$$X_m = \sum_{k=m}^{r} s_k x_k y_k^*, \tag{6.190}$$

and one is to interpret that $X_m = 0$ for $m > r$. Because each operator $\Pi_{m,a}$ is a projection, and the operator $A_a X_m B_a^{\mathsf{T}} \overline{B_a} X_m^* A_a^*$ is positive semidefinite, it follows that

$$\left\langle \Pi_{m,a}, A_a X_m B_a^{\mathsf{T}} \overline{B_a} X_m^* A_a^* \right\rangle \leq \mathrm{Tr}\Big(A_a X_m B_a^{\mathsf{T}} \overline{B_a} X_m^* A_a^* \Big). \tag{6.191}$$

Using the fact that Φ is a channel, and therefore preserves trace, one finds that

$$\begin{aligned} \sum_{a \in \Sigma} & \mathrm{Tr}\Big(A_a X_m B_a^{\mathsf{T}} \overline{B_a} X_m^* A_a^* \Big) \\ &= \mathrm{Tr}\big(\Phi(\mathrm{vec}(X_m)\,\mathrm{vec}(X_m)^*) \big) \\ &= \mathrm{Tr}\big(\mathrm{vec}(X_m)\,\mathrm{vec}(X_m)^* \big) \\ &= \mathrm{Tr}(X_m X_m^*) \\ &= \sum_{k=m}^{n} \lambda_k(XX^*) \end{aligned} \tag{6.192}$$

for each $m \in \{1, \ldots, n\}$.

Finally, as it necessarily holds that $\mathrm{rank}(\Pi_{a,m}) \geq n - m + 1$ for every $a \in \Sigma$ and $m \in \{1, \ldots, n\}$, it follows that

$$\left\langle \Pi_{m,a}, A_a X B_a^{\mathsf{T}} \overline{B_a} X^* A_a^* \right\rangle \geq \sum_{k=m}^{n} \lambda_k\big(A_a X B_a^{\mathsf{T}} \overline{B_a} X^* A_a^* \big). \tag{6.193}$$

By combining (6.188), (6.189), (6.191), (6.192), and (6.193), one finds that

$$\sum_{k=m}^{n} \lambda_k\big(YY^*\big) \leq \sum_{k=m}^{n} \lambda_k(XX^*), \tag{6.194}$$

which establishes (6.180), and therefore completes the proof. □

Theorem 6.37 implies the following corollary, characterizing the pure state transformations from one tensor product space to a possibly different tensor product space that may be realized by LOCC channels.

Corollary 6.38. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be complex Euclidean spaces and let $x \in \mathcal{X} \otimes \mathcal{Y}$ and $y \in \mathcal{Z} \otimes \mathcal{W}$ be unit vectors. The following statements are equivalent:*

1. *For $\rho = \mathrm{Tr}_{\mathcal{Y}}(xx^*)$, $\sigma = \mathrm{Tr}_{\mathcal{W}}(yy^*)$, and $r = \min\{\mathrm{rank}(\rho), \mathrm{rank}(\sigma)\}$, it holds that*
$$\lambda_1(\rho) + \cdots + \lambda_m(\rho) \leq \lambda_1(\sigma) + \cdots + \lambda_m(\sigma) \tag{6.195}$$
for every $m \in \{1, \ldots, r\}$.
2. *There exists a one-way right LOCC channel $\Phi \in \mathrm{LOCC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$ for which it holds that $\Phi(xx^*) = yy^*$.*
3. *There exists a one-way left LOCC channel $\Phi \in \mathrm{LOCC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$ for which it holds that $\Phi(xx^*) = yy^*$.*
4. *There exists a separable channel $\Phi \in \mathrm{SepC}(\mathcal{X}, \mathcal{Z} : \mathcal{Y}, \mathcal{W})$ for which it holds that $\Phi(xx^*) = yy^*$.*

Proof. Define four linear isometries, $A_0 \in \mathrm{U}(\mathcal{X}, \mathcal{X} \oplus \mathcal{Z})$, $B_0 \in \mathrm{U}(\mathcal{Y}, \mathcal{Y} \oplus \mathcal{W})$, $A_1 \in \mathrm{U}(\mathcal{Z}, \mathcal{X} \oplus \mathcal{Z})$, and $B_1 \in \mathrm{U}(\mathcal{W}, \mathcal{Y} \oplus \mathcal{W})$, as follows:

$$\begin{aligned} A_0 x &= x \oplus 0, \quad & A_1 z &= 0 \oplus z, \\ B_0 y &= y \oplus 0, \quad & B_1 w &= 0 \oplus w, \end{aligned} \tag{6.196}$$

for every choice of vectors $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, and $w \in \mathcal{W}$. Also define four channels, $\Psi_0 \in \mathrm{C}(\mathcal{X} \oplus \mathcal{Z}, \mathcal{X})$, $\Lambda_0 \in \mathrm{C}(\mathcal{Y} \oplus \mathcal{W}, \mathcal{Y})$, $\Psi_1 \in \mathrm{C}(\mathcal{X} \oplus \mathcal{Z}, \mathcal{Z})$, and $\Lambda_1 \in \mathrm{C}(\mathcal{Y} \oplus \mathcal{W}, \mathcal{W})$, as

$$\begin{aligned} \Psi_0(X) &= A_0^* X A_0 + \left\langle \mathbb{1}_{\mathcal{X}\oplus\mathcal{Z}} - A_0 A_0^*, X \right\rangle \tau_0, \\ \Lambda_0(Y) &= B_0^* X B_0 + \left\langle \mathbb{1}_{\mathcal{X}\oplus\mathcal{Z}} - B_0 B_0^*, X \right\rangle \xi_0, \\ \Psi_1(X) &= A_1^* X A_1 + \left\langle \mathbb{1}_{\mathcal{X}\oplus\mathcal{Z}} - A_1 A_1^*, X \right\rangle \tau_1, \\ \Lambda_1(Y) &= B_1^* X B_1 + \left\langle \mathbb{1}_{\mathcal{X}\oplus\mathcal{Z}} - B_1 B_1^*, X \right\rangle \xi_1, \end{aligned} \tag{6.197}$$

for all $X \in \mathrm{L}(\mathcal{X} \oplus \mathcal{Z})$ and $Y \in \mathrm{L}(\mathcal{Y} \oplus \mathcal{W})$, where $\tau_0 \in \mathrm{D}(\mathcal{X})$, $\xi_0 \in \mathrm{D}(\mathcal{Y})$, $\tau_1 \in \mathrm{D}(\mathcal{Z})$, and $\xi_1 \in \mathrm{D}(\mathcal{W})$ are fixed, but otherwise arbitrarily selected, density operators.

Assume first that statement 1 holds. One concludes that

$$A_0 \rho A_0^* \prec A_1 \sigma A_1^*, \tag{6.198}$$

and therefore the four equivalent statements of Theorem 6.37 hold for the vectors

$$u = (A_0 \otimes B_0)x \quad \text{and} \quad v = (A_1 \otimes B_1)y. \tag{6.199}$$

There must therefore exist a one-way right LOCC channel Ξ, of the form specified in the statement of Theorem 6.37, such that $\Xi(uu^*) = vv^*$. Define $\Phi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{Y}, \mathcal{Z} \otimes \mathcal{W})$ as

$$\Phi(X) = \big((\Psi_1 \otimes \Lambda_1)\Xi\big)\big((A_0 \otimes B_0)X(A_0 \otimes B_0)^*\big) \tag{6.200}$$

for every $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$. It holds that Φ is a one-way right LOCC channel satisfying $\Phi(xx^*) = yy^*$, and therefore statement 1 implies statement 2. The fact that statement 1 implies statement 3 is similar.

Statements 2 and 3 trivially imply that statement 4 holds.

Finally, assume statement 4 holds. Define a channel Ξ as

$$\Xi(Z) = (A_1 \otimes B_1)\big(\Phi(\Psi_0 \otimes \Lambda_0)\big)(X)(A_1 \otimes B_1)^* \tag{6.201}$$

for all $X \in \mathrm{L}((\mathcal{X} \oplus \mathcal{Z}) \otimes (\mathcal{Y} \oplus \mathcal{W}))$. The channel Ξ is separable and satisfies

$$\Xi(uu^*) = vv^* \tag{6.202}$$

for vectors u and v as in (6.199). The four equivalent statements listed in Theorem 6.37 therefore hold for u and v, which implies

$$\begin{aligned} &\mathrm{Tr}_{\mathcal{Y}\oplus\mathcal{W}}\big((A_0 \otimes B_0)xx^*(A_0 \otimes B_0)^*\big) \\ &\qquad \prec \mathrm{Tr}_{\mathcal{Y}\oplus\mathcal{W}}\big((A_1 \otimes B_1)yy^*(A_1 \otimes B_1)^*\big). \end{aligned} \tag{6.203}$$

This relation is equivalent to

$$A_0 \rho A_0^* \prec A_1 \sigma A_1^*, \tag{6.204}$$

which implies that statement 1 holds, and completes the proof. □

6.2.2 Distillable entanglement and entanglement cost

Suppose $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ is a state, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. There are various ways in which one may quantify the amount of entanglement that is present in ρ, with respect to the bipartition between $\mathcal{X}$ and $\mathcal{Y}$. The *distillable entanglement* and *entanglement cost* represent two such measures. The distillable entanglement concerns the rate at which copies of the state ρ can be converted into copies of the maximally entangled two-qubit state

$$\tau = \frac{1}{2} \sum_{a,b \in \{0,1\}} E_{a,b} \otimes E_{a,b} \tag{6.205}$$

with high accuracy by means of an LOCC channel. The entanglement cost refers to the reverse process; it is the rate at which approximate copies of ρ may be produced from copies of τ by an LOCC channel. In both cases, it is the asymptotic behavior of these processes, as the number of copies of each state grows, that is taken as the measure of entanglement.

For every bipartite state, the distillable entanglement is upper-bounded by the entanglement cost, with the two measures coinciding for pure states. In general, however, the two quantities may differ, with the entanglement cost being strictly larger than the distillable entanglement in some cases.

Notation related to distillable entanglement and entanglement cost

The following notation will be useful when discussing both the distillable entanglement and entanglement cost of a bipartite state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$.

For each $n \in \mathbb{N}$, which represents a number of copies of ρ that are to be manipulated for either of the two measures, complex Euclidean spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$ and $\mathcal{Y}_1, \ldots, \mathcal{Y}_n$ will represent isomorphic copies of $\mathcal{X}$ and $\mathcal{Y}$, respectively, with respect to which the individual copies of ρ are assumed to be defined. Thus, n distinct copies of the state ρ may be represented by the operator

$$\rho^{\otimes n} \in \mathrm{D}\big((\mathcal{X}_1 \otimes \mathcal{Y}_1) \otimes \cdots \otimes (\mathcal{X}_n \otimes \mathcal{Y}_n)\big). \tag{6.206}$$

When it is convenient, the notations

$$\mathcal{X}^{\otimes n} = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \quad \text{and} \quad \mathcal{Y}^{\otimes n} = \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n \tag{6.207}$$

will be used, as well as

$$(\mathcal{X} \otimes \mathcal{Y})^{\otimes n} = (\mathcal{X}_1 \otimes \mathcal{Y}_1) \otimes \cdots \otimes (\mathcal{X}_n \otimes \mathcal{Y}_n). \tag{6.208}$$

One may define an isometry

$$U_n \in \mathrm{U}\big((\mathcal{X} \otimes \mathcal{Y})^{\otimes n}, \mathcal{X}^{\otimes n} \otimes \mathcal{Y}^{\otimes n}\big) \tag{6.209}$$

by the action

$$U_n\big(\mathrm{vec}(A_1) \otimes \cdots \otimes \mathrm{vec}(A_n)\big) = \mathrm{vec}(A_1 \otimes \cdots \otimes A_n) \tag{6.210}$$

for all operators $A_1, \ldots, A_n \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$. Equivalently, U_n is defined by the action

$$\begin{aligned} &U_n\big((x_1 \otimes y_1) \otimes \cdots \otimes (x_n \otimes y_n)\big) \\ &\quad = (x_1 \otimes \cdots \otimes x_n) \otimes (y_1 \otimes \cdots \otimes y_n) \end{aligned} \tag{6.211}$$

for all vectors $x_1, \ldots, x_n \in \mathcal{X}$ and $y_1, \ldots, y_n \in \mathcal{Y}$. This isometry has the effect of re-ordering the tensor factors of the space $(\mathcal{X} \otimes \mathcal{Y})^{\otimes n}$ so that it takes the form of a bipartite tensor product space $\mathcal{X}^{\otimes n} \otimes \mathcal{Y}^{\otimes n}$ that allows for notions concerning entanglement and separability to be conveniently stated .

The binary alphabet will be denoted $\Gamma = \{0, 1\}$, and the state τ defined above is to be considered as an element of the set $\mathrm{D}(\mathcal{Z} \otimes \mathcal{W})$, for $\mathcal{Z} = \mathbb{C}^{\Gamma}$ and $\mathcal{W} = \mathbb{C}^{\Gamma}$. Along the same lines as the convention described above, complex Euclidean spaces $\mathcal{Z}_1, \ldots, \mathcal{Z}_m$ and $\mathcal{W}_1, \ldots, \mathcal{W}_m$, for each $m \in \mathbb{N}$, will denote isomorphic copies of $\mathcal{Z}$ and $\mathcal{W}$, over which distinct copies of the state τ are defined.

Also similar to above, one may define an isometry

$$V_m \in \mathrm{U}\big((\mathcal{Z} \otimes \mathcal{W})^{\otimes m}, \mathcal{Z}^{\otimes m} \otimes \mathcal{W}^{\otimes m}\big) \tag{6.212}$$

playing an analogous role to the isometry U_n, but for the spaces $\mathcal{Z}_1, \ldots, \mathcal{Z}_m$ and $\mathcal{W}_1, \ldots, \mathcal{W}_m$. This isometry is defined by the action

$$V_m\big(\mathrm{vec}(B_1) \otimes \cdots \otimes \mathrm{vec}(B_m)\big) = \mathrm{vec}(B_1 \otimes \cdots \otimes B_m) \tag{6.213}$$

for all operators $B_1, \ldots, B_m \in \mathrm{L}(\mathcal{W}, \mathcal{Z})$. Equivalently, V_m is defined by the action

$$\begin{aligned} &V_m\big((z_1 \otimes w_1) \otimes \cdots \otimes (z_m \otimes w_m)\big) \\ &\quad = (z_1 \otimes \cdots \otimes z_m) \otimes (w_1 \otimes \cdots \otimes w_m) \end{aligned} \tag{6.214}$$

for all vectors $z_1, \ldots, z_m \in \mathcal{Z}$ and $w_1, \ldots, w_m \in \mathcal{W}$.

Definitions of distillable entanglement and entanglement cost

With respect to the notation introduced above, the *distillable entanglement* and *entanglement cost* are defined as follows.

Definition 6.39. Let X and Y be registers with associated complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be a state of (X, Y). With respect to the state ρ, the *distillable entanglement* $\mathrm{E}_{\mathrm{D}}(\mathsf{X} : \mathsf{Y})$ of the pair (X, Y) is defined to be the supremum over all nonnegative real numbers $\alpha \geq 0$ for which the following statement holds: there exists a sequence of LOCC channels $\{\Psi_n : n \in \mathbb{N}\}$, where

$$\Psi_n \in \mathrm{LOCC}\big(\mathcal{X}^{\otimes n}, \mathcal{Z}^{\otimes m} : \mathcal{Y}^{\otimes n}, \mathcal{W}^{\otimes m}\big) \tag{6.215}$$

for $m = \lfloor \alpha n \rfloor$, such that

$$\lim_{n\to\infty} \mathrm{F}\Big(V_m \tau^{\otimes m} V_m^*, \Psi_n\big(U_n \rho^{\otimes n} U_n^*\big)\Big) = 1. \tag{6.216}$$

Definition 6.40. Let X and Y be registers with associated complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be a state of (X, Y). With respect to the state ρ, the *entanglement cost* $\mathrm{E}_{\mathrm{C}}(\mathsf{X} : \mathsf{Y})$ of the pair (X, Y) is defined to be the infimum over all nonnegative real numbers $\alpha \geq 0$ for which the following statement holds: there exists a sequence of LOCC channels $\{\Phi_n : n \in \mathbb{N}\}$, where

$$\Phi_n \in \mathrm{LOCC}\big(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n} : \mathcal{W}^{\otimes m}, \mathcal{Y}^{\otimes n}\big) \tag{6.217}$$

for $m = \lfloor \alpha n \rfloor$, such that

$$\lim_{n\to\infty} \mathrm{F}\Big(U_n \rho^{\otimes n} U_n^*, \Phi_n\big(V_m \tau^{\otimes m} V_m^*\big)\Big) = 1. \tag{6.218}$$

It is intuitive that the entanglement cost should be at least as large as the distillable entanglement, for any choice of $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$, for otherwise one could repeatedly distill copies of the state τ from copies of a given state ρ, use them to produce more copies of ρ, and repeat this process indefinitely, eventually producing any desired number of copies of τ from a finite number of copies of ρ. Such an "entanglement factory" must surely not be possible through local operations and classical communication alone. The following proposition confirms this intuition.

Proposition 6.41. *Let* X *and* Y *be registers. With respect to every state of the pair* (X, Y) *it holds that* $\mathrm{E}_{\mathrm{D}}(\mathsf{X} : \mathsf{Y}) \leq \mathrm{E}_{\mathrm{C}}(\mathsf{X} : \mathsf{Y})$.

Proof. Suppose that n, m, and k are nonnegative integers with $k > m$, and

$$\begin{aligned} \Phi_n &\in \mathrm{LOCC}\big(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n} : \mathcal{W}^{\otimes m}, \mathcal{Y}^{\otimes n}\big) \\ \Psi_n &\in \mathrm{LOCC}\big(\mathcal{X}^{\otimes n}, \mathcal{Z}^{\otimes k} : \mathcal{Y}^{\otimes n}, \mathcal{W}^{\otimes k}\big) \end{aligned} \tag{6.219}$$

are LOCC channels. It holds that

$$V_m \tau^{\otimes m} V_m^* \in \mathrm{Ent}_{2^m}\big(\mathcal{Z}^{\otimes m} : \mathcal{W}^{\otimes m}\big), \tag{6.220}$$

and therefore, given that the composition $\Psi_n \Phi_n$ is an LOCC (and therefore separable) channel, Theorem 6.25 implies that

$$(\Psi_n \Phi_n)\big(V_m \tau^{\otimes m} V_m^*\big) \in \mathrm{Ent}_{2^m}\big(\mathcal{Z}^{\otimes k} : \mathcal{W}^{\otimes k}\big). \tag{6.221}$$

It follows by Proposition 6.16 that

$$\begin{aligned} &\mathrm{F}\Big((\Psi_n \Phi_n)\big(V_m \tau^{\otimes m} V_m^*\big), V_k \tau^{\otimes k} V_k^*\Big)^2 \\ &\quad = \Big\langle (\Psi_n \Phi_n)\big(V_m \tau^{\otimes m} V_m^*\big), V_k \tau^{\otimes k} V_k^* \Big\rangle \leq 2^{m-k} \leq \frac{1}{2}. \end{aligned} \tag{6.222}$$

Now, let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be any state of the pair (X, Y), and suppose α and β are nonnegative real numbers satisfying the definitions of entanglement cost and distillable entanglement, respectively, for the state ρ. For all $\varepsilon > 0$, there must therefore exist a sufficiently large value of $n \in \mathbb{N}$ such that, for $m = \lfloor \alpha n \rfloor$ and $k = \lfloor \beta m \rfloor$, there exist LOCC channels of the form (6.219) for which the following bounds hold:

$$\begin{aligned} \mathrm{F}\Big(\Phi_n\big(V_m \tau^{\otimes m} V_m^*\big), U_n \rho^{\otimes n} U_n^*\Big) &> 1 - \varepsilon, \\ \mathrm{F}\Big(\Psi_n\big(U_n \rho^{\otimes n} V_n^*\big), V_k \tau^{\otimes k} V_k^*\Big) &> 1 - \varepsilon. \end{aligned} \tag{6.223}$$

Therefore, by Theorem 3.32, one may conclude that

$$\mathrm{F}\Big((\Psi_n \Phi_n)\big(V_m \tau^{\otimes m} V_m^*\big), V_k \tau^{\otimes k} V_k^*\Big) > 1 - 4\varepsilon. \tag{6.224}$$

Taking $\varepsilon < 1/8$, one finds from the bound above that $k \leq m$ (for sufficiently large n), and therefore $\alpha \geq \beta$, from it follows that $\mathrm{E}_{\mathrm{D}}(\mathsf{X} : \mathsf{Y}) \leq \mathrm{E}_{\mathrm{C}}(\mathsf{X} : \mathsf{Y})$. □

Pure state entanglement

The next theorem demonstrates that the entanglement cost and distillable entanglement are equal for bipartite pure states; in both cases, the value of these measures agrees with the von Neumann entropy of the states obtained by restricting the given pure state to either part of its bipartition.

Theorem 6.42. *Let* X *and* Y *be registers. With respect to every pure state of the pair* (X, Y)*, one has*

$$\mathrm{E_D}(\mathsf{X}:\mathsf{Y}) = \mathrm{H}(\mathsf{X}) = \mathrm{H}(\mathsf{Y}) = \mathrm{E_C}(\mathsf{X}:\mathsf{Y}). \tag{6.225}$$

Proof. Let $u \in \mathcal{X} \otimes \mathcal{Y}$ be a unit vector, and consider the pure state uu^* of the pair (X, Y). By means of the Schmidt decomposition, one may write

$$u = \sum_{a \in \Sigma} \sqrt{p(a)}\, x_a \otimes y_a \tag{6.226}$$

for some choice of an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and two orthonormal collections $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ and $\{y_a : a \in \Sigma\} \subset \mathcal{Y}$. It holds that

$$\mathrm{Tr}_{\mathcal{Y}}(uu^*) = \sum_{a \in \Sigma} p(a) x_a x_a^* \quad \text{and} \quad \mathrm{Tr}_{\mathcal{X}}(uu^*) = \sum_{a \in \Sigma} p(a) y_a y_a^*, \tag{6.227}$$

which implies that $\mathrm{H}(\mathsf{X}) = \mathrm{H}(p) = \mathrm{H}(\mathsf{Y})$.

Next, recall that, for every choice of $\varepsilon > 0$ and $n \in \mathbb{N}$, the set of ε-typical strings $T_{n,\varepsilon}$ with respect to p contains those strings $a_1 \cdots a_n \in \Sigma^n$ for which

$$2^{-n(\mathrm{H}(p)+\varepsilon)} < p(a_1) \cdots p(a_n) < 2^{-n(\mathrm{H}(p)-\varepsilon)}. \tag{6.228}$$

With this set in mind, one may define a vector $v_{n,\varepsilon} \in \mathcal{X}^{\otimes n} \otimes \mathcal{Y}^{\otimes n}$, for every choice of $\varepsilon > 0$ and $n \in \mathbb{N}$, as

$$v_{n,\varepsilon} = \sum_{a_1 \cdots a_n \in T_{n,\varepsilon}} \sqrt{p(a_1) \cdots p(a_n)}\, x_{a_1 \cdots a_n} \otimes y_{a_1 \cdots a_n}, \tag{6.229}$$

where the shorthand notations

$$x_{a_1 \cdots a_n} = x_{a_1} \otimes \cdots \otimes x_{a_n} \quad \text{and} \quad y_{a_1 \cdots a_n} = y_{a_1} \otimes \cdots \otimes y_{a_n} \tag{6.230}$$

have been used for the sake of brevity. Also define a normalized version of the vector $v_{n,\varepsilon}$ as

$$w_{n,\varepsilon} = \frac{v_{n,\varepsilon}}{\| v_{n,\varepsilon} \|}. \tag{6.231}$$

Observe that

$$2^{-n(\mathrm{H}(p)+\varepsilon)} < \lambda_k\Big(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\big(v_{n,\varepsilon}v_{n,\varepsilon}^*\big)\Big) < 2^{-n(\mathrm{H}(p)-\varepsilon)}, \tag{6.232}$$

and therefore

$$\frac{2^{-n(\mathrm{H}(p)+\varepsilon)}}{\|v_{n,\varepsilon}\|^2} < \lambda_k\Big(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\big(w_{n,\varepsilon}w_{n,\varepsilon}^*\big)\Big) < \frac{2^{-n(\mathrm{H}(p)-\varepsilon)}}{\|v_{n,\varepsilon}\|^2}, \tag{6.233}$$

for $k = 1, \ldots, |T_{n,\varepsilon}|$, while the remaining eigenvalues are zero in both cases.

Now, consider the entanglement cost of the pair (X, Y) with respect to the state uu^*. Let α be any real number such that $\alpha > \mathrm{H}(p)$, let $\varepsilon > 0$ be sufficiently small so that $\alpha > \mathrm{H}(p) + 2\varepsilon$, and consider any choice of $n > 1/\varepsilon$. Denoting $m = \lfloor \alpha n \rfloor$, it follows that $m \geq n(\mathrm{H}(p) + \varepsilon)$. Moreover, it holds that

$$\lambda_k\Big(\mathrm{Tr}_{\mathcal{W}^{\otimes m}}\big(V_m\tau^{\otimes m}V_m^*\big)\Big) = 2^{-m} \tag{6.234}$$

for $k = 1, \ldots, 2^m$. As

$$2^{-m} \leq 2^{-n(\mathrm{H}(p)+\varepsilon)} \leq \frac{2^{-n(\mathrm{H}(p)+\varepsilon)}}{\|v_{n,\varepsilon}\|^2}, \tag{6.235}$$

it follows that

$$\sum_{j=1}^{k} \lambda_j\Big(\mathrm{Tr}_{\mathcal{W}^{\otimes m}}\big(V_m\tau^{\otimes m}V_m^*\big)\Big) \leq \sum_{j=1}^{k} \lambda_j\Big(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\big(w_{n,\varepsilon}w_{n,\varepsilon}^*\big)\Big) \tag{6.236}$$

for every $k \in \{1, \ldots, 2^m\}$. It follows by Corollary 6.38 to Nielsen's theorem (Theorem 6.37) that there exists an LOCC channel

$$\Phi_n \in \mathrm{LOCC}\big(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n} : \mathcal{W}^{\otimes m}, \mathcal{Y}^{\otimes n}\big) \tag{6.237}$$

such that

$$\Phi_n(V_m\tau^{\otimes m}V_m^*) = w_{n,\varepsilon}w_{n,\varepsilon}^*. \tag{6.238}$$

As

$$\mathrm{F}\Big(U_n(uu^*)^{\otimes n}U_n^*, w_{n,\varepsilon}w_{n,\varepsilon}^*\Big)^2 = \sum_{a_1\cdots a_n \in T_{n,\varepsilon}} p(a_1)\cdots p(a_n), \tag{6.239}$$

which approaches 1 in the limit as n approaches infinity, it follows that $\mathrm{E_C}(\mathsf{X}:\mathsf{Y}) \leq \alpha$, and therefore $\mathrm{E_C}(\mathsf{X}:\mathsf{Y}) \leq \mathrm{H}(p)$.

Next, consider the distillable entanglement of (X,Y) with respect to the state uu^*. If $\mathrm{H}(p) = 0$, then there is nothing to prove, as the distillable entanglement is trivially nonnegative, so it will be assumed hereafter that $\mathrm{H}(p) > 0$. Let α be a real number such that $\alpha < \mathrm{H}(p)$, and let $\varepsilon \in (0,1)$ be sufficiently small so that $\alpha < \mathrm{H}(p) - 2\varepsilon$. For all but finitely many values of $n \in \mathbb{N}$, one has that $-\varepsilon n < \log(1-\varepsilon)$, from which it follows that

$$m = \lfloor \alpha n \rfloor \leq n(\mathrm{H}(p) - \varepsilon) + \log(1-\varepsilon), \tag{6.240}$$

and therefore

$$\frac{2^{-n(\mathrm{H}(p)-\varepsilon)}}{1-\varepsilon} \leq 2^{-m}. \tag{6.241}$$

As the quantity

$$\|v_{n,\varepsilon}\|^2 = \sum_{a_1\cdots a_n \in T_{n,\varepsilon}} p(a_1)\cdots p(a_n) \tag{6.242}$$

approaches 1 in the limit as n approaches infinity, it follows that

$$\frac{2^{-n(\mathrm{H}(p)-\varepsilon)}}{\|v_{n,\varepsilon}\|^2} \leq 2^{-m} \tag{6.243}$$

for all but finitely many choices of $n \in \mathbb{N}$.

Now, consider any choice of n for which (6.243) holds (where $m = \lfloor \alpha n \rfloor$ as usual). One therefore has

$$\sum_{j=1}^{k} \lambda_j\Big(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\big(w_{n,\varepsilon} w_{n,\varepsilon}^*\big)\Big) \leq \sum_{j=1}^{k} \lambda_j\Big(\mathrm{Tr}_{\mathcal{W}^{\otimes m}}\big(V_m \tau^{\otimes m} V_m^*\big)\Big) \tag{6.244}$$

for every $k \in \{1,\ldots,2^m\}$. Again using Corollary 6.38, one has that there must exist an LOCC channel

$$\Phi_n \in \mathrm{LOCC}\big(\mathcal{X}^{\otimes n}, \mathcal{Z}^{\otimes m} : \mathcal{Y}^{\otimes n}, \mathcal{W}^{\otimes m}\big) \tag{6.245}$$

such that

$$\Phi_n\big(w_{n,\varepsilon} w_{n,\varepsilon}^*\big) = V_m \tau^{\otimes m} V_m^*. \tag{6.246}$$

Making use of the monotonicity of the fidelity function under the action of

a channel (Theorem 3.30), one finds that

$$\begin{aligned}
&\mathrm{F}\Big(\Phi_n\big(U_n(uu^*)^{\otimes n}U_n^*\big), V_m\tau^{\otimes m}V_m^*\Big)^2 \\
&\qquad = \mathrm{F}\Big(\Phi_n\big(U_n(uu^*)^{\otimes n}U_n^*\big), \Phi_n(w_{n,\varepsilon}w_{n,\varepsilon}^*)\Big)^2 \\
&\qquad \geq \mathrm{F}\Big(U_n(uu^*)^{\otimes n}U_n^*\,,\, w_{n,\varepsilon}w_{n,\varepsilon}^*\Big)^2 \\
&\qquad = \sum_{a_1\cdots a_n\in T_{n,\varepsilon}} p(a_1)\cdots p(a_n).
\end{aligned} \tag{6.247}$$

The quantity on the right-hand-side of this inequality approaches 1 in the limit as n approaches infinity, from which it follows that $\mathrm{E}_{\mathrm{D}}(\mathsf{X}:\mathsf{Y}) \geq \alpha$, and therefore $\mathrm{E}_{\mathrm{D}}(\mathsf{X}:\mathsf{Y}) \geq \mathrm{H}(p)$.

It has been proved that

$$\mathrm{E}_{\mathrm{C}}(\mathsf{X}:\mathsf{Y}) \leq \mathrm{H}(p) \leq \mathrm{E}_{\mathrm{D}}(\mathsf{X}:\mathsf{Y}). \tag{6.248}$$

The inequality $\mathrm{E}_{\mathrm{D}}(\mathsf{X}:\mathsf{Y}) \leq \mathrm{E}_{\mathrm{C}}(\mathsf{X}:\mathsf{Y})$ holds by Proposition 6.41, so the proof is complete. □

Remark 6.43. For a given unit vector $u \in \mathcal{X}\otimes\mathcal{Y}$, for complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, the quantity in (6.225) is known as the *entanglement entropy* of the pure state uu^*.

6.2.3 Bound entanglement and partial transposition

Informally speaking, Theorem 6.42 implies that all pure state entanglement is equivalent in the bipartite setting. A bipartite pure state is entangled if and only if it has positive entanglement entropy. Moreover, given any two entangled pure states, one necessarily has that an approximate conversion between a large number of copies of the first state to the second state is possible through the use of an LOCC channel, at a rate determined by the ratio of the entanglement entropies of the two states.

The situation is more complex for mixed states. One respect in which this is so is that there exist entangled states having no distillable entanglement. The entanglement in such states, which is referred to as *bound entanglement*, can never be converted into pure state entanglement through the use of an LOCC channel. The fact that states of this sort exist may be proved through the use of properties of the transpose mapping.

The partial transpose and separability

For any complex Euclidean space $\mathcal{X}$, the transpose mapping $\mathrm{T} \in \mathrm{T}(\mathcal{X})$ is defined as

$$\mathrm{T}(X) = X^{\mathsf{T}} \tag{6.249}$$

for all $X \in \mathrm{L}(\mathcal{X})$. As this is a positive map, it follows by the Horodecki criterion (Theorem 6.10) that

$$\bigl(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\bigr)(R) \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}) \tag{6.250}$$

for every separable operator $R \in \mathrm{Sep}(\mathcal{X} : \mathcal{Y})$. If $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ is a positive semidefinite operator for which

$$\bigl(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\bigr)(P) \notin \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}), \tag{6.251}$$

then one may therefore conclude that P is not separable.

The converse of this statement does not hold in general. Given a positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ for which

$$\bigl(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\bigr)(P) \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}), \tag{6.252}$$

one may not conclude that P is separable; an example of a non-separable operator possessing the property (6.252) is described below.

It is the case, however, that an operator $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ satisfying the condition (6.252) is highly constrained, in some sense, with respect to the way it is entangled. With this idea in mind, one defines the sets of PPT operators and PPT states (short for *positive partial transpose* operators and states) as follows.

Definition 6.44. For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, the set $\mathrm{PPT}(\mathcal{X} : \mathcal{Y})$ is defined as the set of all operators $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ that satisfy

$$\bigl(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\bigr)(P) \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}). \tag{6.253}$$

Elements of the set $\mathrm{PPT}(\mathcal{X} : \mathcal{Y})$ are called *PPT operators*, while elements of the set $\mathrm{PPT}(\mathcal{X} : \mathcal{Y}) \cap \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ are called *PPT states*.

Unextendable product sets and non-separable PPT operators

One method by which non-separable PPT operators may be constructed involves the notion of an *unextendable product set*. For complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, an orthonormal collection of vectors of the form

$$\mathcal{A} = \{u_1 \otimes v_1, \ldots, u_m \otimes v_m\}, \tag{6.254}$$

for unit vectors $u_1, \ldots, u_m \in \mathcal{X}$ and $v_1, \ldots, v_m \in \mathcal{Y}$, is an *unextendable product set* if two properties hold:

1. $\mathcal{A}$ spans a proper subspace of $\mathcal{X} \otimes \mathcal{Y}$. (Equivalently, $m < \dim(\mathcal{X} \otimes \mathcal{Y})$.)
2. For every choice of vectors $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ satisfying $x \otimes y \perp \mathcal{A}$, it must hold that $x \otimes y = 0$.

Example 6.45. Define unit vectors $u_1, \ldots, u_5 \in \mathcal{X}$ and $v_1, \ldots, v_5 \in \mathcal{Y}$, for $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_3}$ and $\mathcal{Y} = \mathbb{C}^{\mathbb{Z}_3}$, as follows:

$$\begin{aligned}
u_1 &= e_0\,, & v_1 &= \frac{1}{\sqrt{2}}(e_0 - e_1)\,,\\
u_2 &= e_2\,, & v_2 &= \frac{1}{\sqrt{2}}(e_1 - e_2)\,,\\
u_3 &= \frac{1}{\sqrt{2}}(e_0 - e_1)\,, & v_3 &= e_2\,,\\
u_4 &= \frac{1}{\sqrt{2}}(e_1 - e_2)\,, & v_4 &= e_0\,,\\
u_5 &= \frac{1}{\sqrt{3}}(e_0 + e_1 + e_2)\,, & v_5 &= \frac{1}{\sqrt{3}}(e_0 + e_1 + e_2).
\end{aligned} \tag{6.255}$$

It therefore holds that

$$\begin{aligned}
u_1 \otimes v_1 &= \frac{1}{\sqrt{2}}\, e_0 \otimes (e_0 - e_1),\\
u_2 \otimes v_2 &= \frac{1}{\sqrt{2}}\, e_2 \otimes (e_1 - e_2),\\
u_3 \otimes v_3 &= \frac{1}{\sqrt{2}}\, (e_0 - e_1) \otimes e_2,\\
u_4 \otimes v_4 &= \frac{1}{\sqrt{2}}\, (e_1 - e_2) \otimes e_0,\\
u_5 \otimes v_5 &= \frac{1}{3}\, (e_0 + e_1 + e_2) \otimes (e_0 + e_1 + e_2).
\end{aligned} \tag{6.256}$$

The set $\{u_1 \otimes v_1, \ldots, u_5 \otimes v_5\}$ is orthonormal by inspection. If $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ satisfy

$$\langle x \otimes y, u_k \otimes v_k \rangle = \langle x, u_k \rangle \langle y, v_k \rangle = 0 \tag{6.257}$$

for $k = 1, \ldots, 5$, then one must have $\langle x, u_k \rangle = 0$ for at least 3 distinct choices of $k \in \{1, \ldots, 5\}$ or $\langle y, v_k \rangle = 0$ for at least 3 distinct choices of $k \in \{1, \ldots, 5\}$. As every 3 distinct choices of u_k span all of $\mathcal{X}$ and every 3 distinct choices of v_k span all of $\mathcal{Y}$, it follows that $x \otimes y = 0$. The set $\{u_1 \otimes v_1, \ldots, u_5 \otimes v_5\}$ is therefore an unextendable product set.

The projection onto the subspace orthogonal to an unextendable product set must be both PPT and entangled, as the following theorem states.

Theorem 6.46. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let*

$$\{u_1 \otimes v_1, \ldots, u_m \otimes v_m\} \tag{6.258}$$

be an unextendable product set in $\mathcal{X} \otimes \mathcal{Y}$, and define

$$\Pi = \sum_{k=1}^{m} u_k u_k^* \otimes v_k v_k^*. \tag{6.259}$$

It holds that

$$\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - \Pi \in \mathrm{PPT}(\mathcal{X} : \mathcal{Y}) \setminus \mathrm{Sep}(\mathcal{X} : \mathcal{Y}). \tag{6.260}$$

Proof. From the assumption that $\{u_1 \otimes v_1, \ldots, u_m \otimes v_m\}$ is an orthonormal set, one may conclude that $\{\overline{u_1} \otimes v_1, \ldots, \overline{u_k} \otimes v_k\}$ is an orthonormal set as well. It follows that

$$\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\Pi) = \sum_{k=1}^{m} \overline{u_k} u_k^{\mathsf{T}} \otimes v_k v_k^* \tag{6.261}$$

is a projection operator, and therefore

$$\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\Pi) \leq \mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}}. \tag{6.262}$$

As

$$\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}}) = \mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}}, \tag{6.263}$$

one obtains the inclusion

$$\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - \Pi) \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y}). \tag{6.264}$$

Now, toward a contradiction, assume that

$$\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - \Pi \in \operatorname{Sep}(\mathcal{X} : \mathcal{Y}), \tag{6.265}$$

which implies that

$$\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - \Pi = \sum_{a \in \Sigma} x_a x_a^* \otimes y_a y_a^* \tag{6.266}$$

for some choice of an alphabet Σ and collections $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ and $\{y_a : a \in \Sigma\} \subset \mathcal{Y}$. It holds that

$$\begin{aligned} &\sum_{k=1}^{m} \sum_{a \in \Sigma} |\langle x_a \otimes y_a, u_k \otimes v_k \rangle|^2 \\ &\quad = \sum_{k=1}^{m} (u_k \otimes v_k)^* (\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - \Pi)(u_k \otimes v_k) = 0, \end{aligned} \tag{6.267}$$

and therefore $\langle x_a \otimes y_a, u_k \otimes v_k \rangle = 0$ for every $a \in \Sigma$ and $k \in \{1, \ldots, m\}$. By the assumption that $\{u_1 \otimes v_1, \ldots, u_m \otimes v_m\}$ is an unextendable product set, it follows that $x_a \otimes y_a = 0$ for every $a \in \Sigma$, and therefore

$$\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - \Pi = 0. \tag{6.268}$$

This, however, is in contradiction with the assumption $m < \dim(\mathcal{X} \otimes \mathcal{Y})$. It follows that

$$\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}} - \Pi \notin \operatorname{Sep}(\mathcal{X} : \mathcal{Y}), \tag{6.269}$$

which completes the proof. □

PPT states have no distillable entanglement

PPT states may not always be separable, but they exhibit similar properties to separable states in some respects. One such respect is that their overlap with every maximally entangled state is small. The next proposition, which is reminiscent of Proposition 6.16, is representative of this fact.

Proposition 6.47. *Let $Y \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be an operator satisfying $\|Y\| \leq 1$, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. For every operator $P \in \mathrm{PPT}(\mathcal{X} : \mathcal{Y})$ it holds that*

$$\left\langle \operatorname{vec}(Y) \operatorname{vec}(Y)^*, P \right\rangle \leq \operatorname{Tr}(P). \tag{6.270}$$

Proof. Observe that one may write

$$\operatorname{vec}(Y) = \left(\mathbb{1}_{\mathcal{X}} \otimes Y^{\mathsf{T}}\right) \operatorname{vec}(\mathbb{1}_{\mathcal{X}}), \tag{6.271}$$

which implies that

$$(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\operatorname{vec}(Y) \operatorname{vec}(Y)^*) = \left(\mathbb{1}_{\mathcal{X}} \otimes Y^{\mathsf{T}}\right) W \left(\mathbb{1}_{\mathcal{X}} \otimes Y^{\mathsf{T}}\right)^* \tag{6.272}$$

for $W \in \mathrm{U}(\mathcal{X} \otimes \mathcal{X})$ denoting the swap operator on $\mathcal{X} \otimes \mathcal{X}$. It holds that

$$\begin{aligned} &\big\langle \operatorname{vec}(Y) \operatorname{vec}(Y)^*, P \big\rangle \\ &\quad = \big\langle (\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(\operatorname{vec}(Y) \operatorname{vec}(Y)^*), (\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(P) \big\rangle \\ &\quad \leq \big\| (\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(P) \big\|_1 \\ &\quad = \operatorname{Tr}(P); \end{aligned} \tag{6.273}$$

the first equality follows from the fact that the transpose mapping is its own adjoint and inverse, the inequality follows from the fact that the operator (6.272) has spectral norm at most 1, and the second equality follows from the assumption that $P \in \operatorname{PPT}(\mathcal{X} : \mathcal{Y})$ along with the observation that the transpose mapping preserves trace. □

Example 6.48. Similar to Example 6.17, let Σ be an alphabet, let $n = |\Sigma|$, and let $\mathcal{X} = \mathbb{C}^{\Sigma}$ and $\mathcal{Y} = \mathbb{C}^{\Sigma}$. Define a density operator $\tau \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ as

$$\tau = \frac{1}{n} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b} = \frac{1}{n} \operatorname{vec}(\mathbb{1}) \operatorname{vec}(\mathbb{1})^*, \tag{6.274}$$

where $\mathbb{1}$ denotes the identity operator on $\mathbb{C}^{\Sigma}$, which may be viewed as an element of the set $\mathrm{L}(\mathcal{Y}, \mathcal{X})$. For every PPT density operator

$$\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y}) \cap \operatorname{PPT}(\mathcal{X} : \mathcal{Y}), \tag{6.275}$$

it holds that

$$\langle \tau, \rho \rangle = \frac{1}{n} \big\langle \operatorname{vec}(\mathbb{1}) \operatorname{vec}(\mathbb{1})^*, \rho \big\rangle \leq \frac{1}{n} \tag{6.276}$$

by Proposition 6.47. Thus, with respect to their overlap with the maximally entangled state τ, one has that PPT operators are bounded in a similar way to separable operators.

Proposition 6.47, when combined with the following proposition stating that separable maps (and therefore LOCC channels) map PPT operators to PPT operators, leads to a proof that PPT states have distillable entanglement equal to zero.

Proposition 6.49. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be complex Euclidean spaces, and suppose that $P \in \mathrm{PPT}(\mathcal{X}:\mathcal{Y})$ is a PPT operator and $\Phi \in \mathrm{SepCP}(\mathcal{X},\mathcal{Z}:\mathcal{Y},\mathcal{W})$ is a separable map. It holds that $\Phi(P) \in \mathrm{PPT}(\mathcal{Z}:\mathcal{W})$.*

Proof. For any choice of operators $A \in \mathrm{L}(\mathcal{X},\mathcal{Z})$ and $B \in \mathrm{L}(\mathcal{Y},\mathcal{W})$, it holds that

$$\begin{aligned}&\big(\mathrm{T}\otimes\mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big((A\otimes B)P(A\otimes B)^*\big)\\ &\quad= \big(\overline{A}\otimes B\big)\big(\mathrm{T}\otimes\mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(P)\big(\overline{A}\otimes B\big)^* \in \mathrm{Pos}\big(\mathcal{Z}\otimes\mathcal{W}\big),\end{aligned} \tag{6.277}$$

by virtue of the fact that $P \in \mathrm{PPT}(\mathcal{X}:\mathcal{Y})$. As Φ is separable, one has

$$\Phi(X) = \sum_{a\in\Sigma} (A_a\otimes B_a)X(A_a\otimes B_a)^* \tag{6.278}$$

for all $X \in \mathrm{L}(\mathcal{X}\otimes\mathcal{Y})$, for some choice of an alphabet Σ and collections of operators

$$\{A_a : a\in\Sigma\} \subset \mathrm{L}(\mathcal{X},\mathcal{Z}) \;\text{ and }\; \{B_a : a\in\Sigma\} \subset \mathrm{L}(\mathcal{Y},\mathcal{W}). \tag{6.279}$$

Consequently, one has that

$$\big(\mathrm{T}\otimes\mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\Phi(P)) = \sum_{a\in\Sigma}\big(\overline{A_a}\otimes B_a\big)\big(\mathrm{T}\otimes\mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(P)\big(\overline{A_a}\otimes B_a\big)^* \tag{6.280}$$

is a positive semidefinite operator, and therefore $\Phi(P) \in \mathrm{PPT}\big(\mathcal{Z}:\mathcal{W}\big)$, as required. □

Theorem 6.50. *Let X and Y be registers and consider a PPT state*

$$\rho \in \mathrm{PPT}(\mathcal{X}:\mathcal{Y}) \cap \mathrm{D}(\mathcal{X}\otimes\mathcal{Y}) \tag{6.281}$$

of the pair (X,Y). With respect to the state ρ, it holds that $\mathrm{E}_{\mathrm{D}}(\mathsf{X}:\mathsf{Y}) = 0$.

Proof. Let $\Gamma = \{0,1\}$, let $\mathcal{Z} = \mathbb{C}^{\Gamma}$ and $\mathcal{W} = \mathbb{C}^{\Gamma}$, and let $\tau \in \mathrm{D}(\mathcal{Z}\otimes\mathcal{W})$ be defined as

$$\tau = \frac{1}{2}\sum_{a,b\in\Gamma} E_{a,b}\otimes E_{a,b}. \tag{6.282}$$

Let n and m be arbitrary positive integers, consider any LOCC channel

$$\Phi \in \mathrm{LOCC}\big(\mathcal{X}^{\otimes n},\mathcal{Z}^{\otimes m}:\mathcal{Y}^{\otimes n},\mathcal{W}^{\otimes m}\big), \tag{6.283}$$

and recall the operators U_n and V_m as defined by (6.211) and (6.213). It holds that

$$U_n \rho^{\otimes n} U_n^* \in \mathrm{PPT}\big(\mathcal{X}^{\otimes n} : \mathcal{Y}^{\otimes n}\big), \tag{6.284}$$

and therefore

$$\Phi(V_m \rho^{\otimes n} V_m^*) \in \mathrm{PPT}\big(\mathcal{Z}^{\otimes m} : \mathcal{W}^{\otimes m}\big) \tag{6.285}$$

by Proposition 6.49. One may therefore conclude from Proposition 6.47 that

$$\mathrm{F}\big(V_m \tau^{\otimes m} V_m^*, \Phi(U_n \rho^{\otimes n} U_n)\big) \leq 2^{-\frac{m}{2}} \leq \frac{1}{\sqrt{2}}. \tag{6.286}$$

It follows that $\mathrm{E}_{\mathrm{D}}(\mathsf{X} : \mathsf{Y}) = 0$. □

6.3 Phenomena associated with entanglement

This section discusses two notions—teleportation (together with the related notion of dense coding) and non-classical correlations—that are generally associated with entanglement, and serve as representatives of the sorts of operational effects that entanglement may induce.

6.3.1 Teleportation and dense coding

In the setting of quantum information, *teleportation* has traditionally referred to a protocol by which a single-qubit quantum channel is implemented through the use of a maximally entangled pair of qubits combined with two classical bits of communication. Informally speaking, teleportation suggests the transformation

> 1 pair of maximally entangled qubits
> + 2 bits of classical communication
> $\rightarrow$ 1 qubit of quantum communication.

Teleportation is often associated with the *dense coding* protocol, which offers a complementary trade-off between resources; dense coding traditionally refers to a protocol by which a two-bit classical channel is implemented through the use of a maximally entangled pair of qubits and a single-qubit quantum channel. In this case, the suggested transformation is

1 pair of maximally entangled qubits
+ 1 qubit of quantum communication
$\rightarrow$ 2 bits of classical communication.

In both cases, the maximally entangled pair of qubits is consumed by the conversion between two classical bits and one qubit of communication; in essence, the entangled pair of qubits functions as a resource allowing for this conversion.

In the discussion that follows, teleportation and dense coding will be considered in greater generality. The traditional protocols suggested above will emerge as specific instances of more general classes of protocols.

Teleportation

Consider the following scenario in which two individuals, Alice and Bob, aim to implement an ideal quantum channel through the combined use of entanglement and classical communication.

Scenario 6.51 (Teleportation)**.** Alice holds a register X and Bob holds Y. Both registers have the same classical state set Σ, and the state of the pair (X, Y) is given by the maximally entangled state

$$\tau = \frac{1}{|\Sigma|} \sum_{b,c \in \Sigma} E_{b,c} \otimes E_{b,c}. \tag{6.287}$$

Alice obtains a new register Z, whose classical state set is also Σ, and she wishes to transmit Z to Bob. Alice and Bob attempt to accomplish this task using classical communication together with the shared entangled state τ, by means of a protocol as follows:

1. Alice performs a measurement $\mu : \Gamma \rightarrow \mathrm{Pos}(\mathcal{Z} \otimes \mathcal{X})$ on the pair (Z, X), where Γ is an arbitrarily chosen alphabet, and sends the outcome $a \in \Gamma$ of this measurement to Bob.
2. For $\{\Psi_a : a \in \Gamma\} \subseteq \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ being a collection of channels indexed by Γ, Bob applies the channel Ψ_a to Y, for whichever symbol $a \in \Gamma$ was sent to him by Alice, transforming this register into a new register Z.

An analysis reveals that this protocol accomplishes the task at hand for a suitable choice for Alice's measurement and Bob's collection of channels.

Remark 6.52. One may consider more general scenarios along similar lines to Scenario 6.51. For instance, X, Y, and Z might not share the same classical state set, the initial state of the pair (X, Y) might be initialized to a different state than τ, and Alice and Bob might aim to implement a channel different from the identity channel. The discussion that follows, however, will focus on the setting described in Scenario 6.51 in the interest of simplicity.

For a given choice of Alice's measurement μ and Bob's collection of channels $\{\Psi_a : a \in \Gamma\}$, the channel $\Phi \in \mathrm{C}(\mathcal{Z})$ that is implemented by the protocol described in Scenario 6.51 may be expressed as

$$\Phi(Z) = \frac{1}{|\Sigma|} \sum_{a \in \Gamma} \sum_{b,c \in \Sigma} \langle \mu(a), Z \otimes E_{b,c} \rangle \Psi_a(E_{b,c}) \tag{6.288}$$

for all $Z \in \mathrm{L}(\mathcal{Z})$. The following theorem provides a characterization of those measurements and collections of channels for which the channel Φ is equal to the identity channel, which represents an ideal transmission of quantum information from Alice to Bob. (The statement of the theorem includes the assumption that none of the measurement operators of μ are identically zero, as this allows for a cleaner statement of the characterization.)

Theorem 6.53. *Let Σ and Γ be alphabets, let $\mathcal{X} = \mathbb{C}^{\Sigma}$, $\mathcal{Y} = \mathbb{C}^{\Sigma}$, and $\mathcal{Z} = \mathbb{C}^{\Sigma}$ be complex Euclidean spaces, let $\mu : \Gamma \to \mathrm{Pos}(\mathcal{Z} \otimes \mathcal{X})$ be a measurement such that $\mu(a) \neq 0$ for every $a \in \Gamma$, and let $\{\Psi_a : a \in \Gamma\} \subseteq \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ be a collection of channels. The following two statements are equivalent:*

1. *It holds that*
$$Z = \frac{1}{|\Sigma|} \sum_{a \in \Gamma} \sum_{b,c \in \Sigma} \langle \mu(a), Z \otimes E_{b,c} \rangle \Psi_a(E_{b,c}) \tag{6.289}$$
for every $Z \in \mathrm{L}(\mathcal{Z})$.
2. *There exists a collection $\{U_a : a \in \Gamma\} \subset \mathrm{U}(\mathbb{C}^{\Sigma})$ of unitary operators and a probability vector $p \in \mathcal{P}(\Gamma)$ such that*
$$\mu(a) = p(a)|\Sigma| \operatorname{vec}(U_a) \operatorname{vec}(U_a)^* \quad \text{and} \quad \Psi_a(Y) = U_a Y U_a^* \tag{6.290}$$
for every choice of $a \in \Gamma$ and $Y \in \mathrm{L}(\mathcal{Y})$.

The proof of Theorem 6.53 will make use of the following proposition, which establishes that a channel of the form $\Phi \in \mathrm{C}(\mathcal{X})$, for any complex Euclidean space $\mathcal{X}$, has a completely positive inverse only if Φ is a unitary channel.

Proposition 6.54. *Let $\mathcal{X}$ be a complex Euclidean space, let $\Phi \in \mathrm{C}(\mathcal{X})$ be a channel, and let $\Psi \in \mathrm{CP}(\mathcal{X})$ be a completely positive map for which $\Phi\Psi = \mathbb{1}_{\mathrm{L}(\mathcal{X})}$. There exists a unitary operator $U \in \mathrm{U}(\mathcal{X})$ such that*

$$\Phi(X) = U^* X U \quad \text{and} \quad \Psi(X) = U X U^* \tag{6.291}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

Proof. As Ψ is completely positive, and evidently nonzero, its Choi operator $J(\Psi)$ is a nonzero positive semidefinite operator. By the spectral theorem (Corollary 1.4), it is therefore possible to write

$$J(\Psi) = \sum_{k=1}^{r} \operatorname{vec}(A_k) \operatorname{vec}(A_k)^* \tag{6.292}$$

for $r = \operatorname{rank}(J(\Psi))$ and $\{A_1, \ldots, A_r\} \subset \mathrm{L}(\mathcal{X})$ being an orthogonal collection of nonzero operators. Consequently, one has

$$\begin{aligned} &\sum_{k=1}^{r} \bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\bigr)\bigl(\operatorname{vec}(A_k) \operatorname{vec}(A_k)^*\bigr) \\ &\quad = \bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\bigr)(J(\Psi)) = J(\Phi\Psi) = \operatorname{vec}(\mathbb{1}_{\mathcal{X}}) \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^*. \end{aligned} \tag{6.293}$$

As $\operatorname{vec}(\mathbb{1}_{\mathcal{X}}) \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^*$ has rank equal to one, and each operator

$$\bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\bigr)\bigl(\operatorname{vec}(A_k) \operatorname{vec}(A_k)^*\bigr) \tag{6.294}$$

is positive semidefinite (by the complete positivity of Φ), it follows that there must exist a probability vector $(p_1, \ldots, p_r)$ such that

$$\bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\bigr)\bigl(\operatorname{vec}(A_k) \operatorname{vec}(A_k)^*\bigr) = p_k \operatorname{vec}(\mathbb{1}_{\mathcal{X}}) \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^* \tag{6.295}$$

for each $k \in \{1, \ldots, r\}$. Because Φ preserves trace, it follows that

$$\bigl(A_k^* A_k\bigr)^{\mathsf{T}} = \bigl(\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\bigr)\bigl(\operatorname{vec}(A_k) \operatorname{vec}(A_k)^*\bigr) = p_k \mathbb{1}_{\mathcal{X}}, \tag{6.296}$$

and therefore $A_k = \sqrt{p_k} U_k$ for some choice of a unitary operator $U_k \in \mathrm{U}(\mathcal{X})$, for each $k \in \{1, \ldots, r\}$. This implies that

$$\begin{aligned} \bigl(\mathbb{1}_{\mathcal{X}} \otimes U_k^{\mathsf{T}}\bigr) J(\Phi) \bigl(\mathbb{1}_{\mathcal{X}} \otimes U_k^{\mathsf{T}}\bigr)^* &= \bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\bigr)\bigl(\operatorname{vec}(U_k) \operatorname{vec}(U_k)^*\bigr) \\ &= \operatorname{vec}(\mathbb{1}_{\mathcal{X}}) \operatorname{vec}(\mathbb{1}_{\mathcal{X}})^*, \end{aligned} \tag{6.297}$$

and therefore

$$J(\Phi) = \operatorname{vec}(U_k^*)\operatorname{vec}(U_k^*)^*, \tag{6.298}$$

again for each $k \in \{1, \ldots, r\}$. As $\{A_1, \ldots, A_r\}$ is a collection of nonzero, orthogonal operators, and is therefore linearly independent, one concludes that $r = 1$ and $p_1 = 1$; and by setting $U = U_1$ the proposition is proved. □

Proof of Theorem 6.53. Assume first that statement 1 holds. For each $a \in \Gamma$, define a map $\Xi_a \in \mathrm{T}(\mathbb{C}^\Sigma)$ as

$$\Xi_a(Z) = \frac{1}{|\Sigma|} \sum_{b,c \in \Sigma} \langle \mu(a), Z \otimes E_{b,c} \rangle \, E_{b,c} \tag{6.299}$$

for all $Z \in \mathrm{L}(\mathbb{C}^\Sigma)$. The Choi operator of Ξ_a is given by

$$\mathrm{J}(\Xi_a) = \frac{1}{|\Sigma|} W \overline{\mu(a)} W, \tag{6.300}$$

for $W \in \mathrm{U}(\mathbb{C}^\Sigma \otimes \mathbb{C}^\Sigma)$ denoting the swap operator. As $J(\Xi_a)$ is positive semidefinite for each $a \in \Gamma$, it follows that Ξ_a is completely positive, and moreover is nonzero by the assumption that $\mu(a)$ is nonzero. Statement 1 may now be expressed as

$$\sum_{a \in \Gamma} \Psi_a \Xi_a = \mathbb{1}_{\mathrm{L}(\mathcal{Z})}, \tag{6.301}$$

which is equivalent to

$$\sum_{a \in \Gamma} \mathrm{J}(\Psi_a \Xi_a) = \operatorname{vec}(\mathbb{1}_\mathcal{Z}) \operatorname{vec}(\mathbb{1}_\mathcal{Z})^*. \tag{6.302}$$

As the composition $\Psi_a \Xi_a$ is necessarily completely positive and nonzero for each $a \in \Gamma$, and the operator $\operatorname{vec}(\mathbb{1}_\mathcal{Z}) \operatorname{vec}(\mathbb{1}_\mathcal{Z})^*$ has rank equal to 1, it follows that there must exist a probability vector $p \in \mathcal{P}(\Gamma)$ such that

$$\mathrm{J}(\Psi_a \Xi_a) = p(a) \operatorname{vec}(\mathbb{1}_\mathcal{Z}) \operatorname{vec}(\mathbb{1}_\mathcal{Z})^* \tag{6.303}$$

for each $a \in \Gamma$. Consequently,

$$\frac{(\Psi_a \Xi_a)(Z)}{p(a)} = Z \tag{6.304}$$

for every $Y \in \mathrm{L}(\mathcal{Z})$. By Proposition 6.54, there must exist a collection of unitary operators $\{U_a : a \in \Sigma\} \subset \mathrm{U}(\mathbb{C}^\Sigma)$ such that

$$\Psi_a(Z) = U_a Z U_a^* \quad \text{and} \quad \frac{1}{p(a)}\Xi_a(Z) = U_a^* Z U_a \tag{6.305}$$

for every $a \in \Gamma$ and $Z \in \mathrm{L}(\mathbb{C}^\Sigma)$. Thus,

$$\frac{1}{|\Sigma|} W \overline{\mu(a)} W = \mathrm{J}(\Xi_a) = p(a)\,\mathrm{vec}(U_a^*)\,\mathrm{vec}(U_a^*)^*, \tag{6.306}$$

and because $W\,\mathrm{vec}(A) = \mathrm{vec}(A^\mathsf{T})$ for every $A \in \mathrm{L}(\mathbb{C}^\Sigma)$, one therefore has

$$\mu(a) = p(a)\,|\Sigma|\,\mathrm{vec}(U_a)\,\mathrm{vec}(U_a)^* \tag{6.307}$$

for each $a \in \Gamma$. Statement 1 therefore implies statement 2.

Now assume statement 2 holds. As μ is assumed to be a measurement, it must be the case that

$$|\Sigma| \sum_{a\in\Gamma} p(a)\,\mathrm{vec}(U_a)\,\mathrm{vec}(U_a)^* = \mathbb{1}_\Sigma \otimes \mathbb{1}_\Sigma, \tag{6.308}$$

and therefore

$$\sum_{a\in\Gamma} p(a)\,\mathrm{vec}(U_a)\,\mathrm{vec}(U_a)^* = \frac{1}{|\Sigma|}\mathbb{1}_\Sigma \otimes \mathbb{1}_\Sigma. \tag{6.309}$$

The operator represented by the equation (6.309) coincides with the Choi operator $\mathrm{J}(\Omega)$ of the completely depolarizing channel $\Omega \in \mathrm{C}(\mathbb{C}^\Sigma)$. It follows that one may write

$$\Omega(X) = \sum_{a\in\Gamma} p(a) U_a X U_a^* \tag{6.310}$$

for every $X \in \mathrm{L}(\mathbb{C}^\Sigma)$. As the natural representation $\mathrm{K}(\Omega)$ of the completely depolarizing channel is equal to the operator τ, one has that

$$\sum_{a\in\Gamma} p(a)\, U_a \otimes \overline{U_a} = \mathrm{K}(\Omega) = \tau \tag{6.311}$$

by Proposition 2.20, and because τ is invariant under taking the entry-wise complex conjugate it follows that

$$\sum_{a\in\Gamma} p(a)\, \overline{U_a} \otimes U_a = \tau. \tag{6.312}$$

Now consider the channel $\Phi \in \mathrm{C}(\mathbb{C}^{\Sigma})$ defined by

$$\Phi(Z) = \frac{1}{|\Sigma|} \sum_{a \in \Gamma} \sum_{b,c \in \Sigma} \langle \mu(a), Z \otimes E_{b,c} \rangle \Psi_a(E_{b,c}) \tag{6.313}$$

for every $Z \in \mathrm{L}(\mathcal{Z})$. Making use of the expression (6.312), one may write

$$\Phi(Z) = \sum_{a,b \in \Gamma} p(b) \big\langle \mu(a), Z \otimes \overline{U_b} \big\rangle \Psi_a(U_b) \tag{6.314}$$

for every $Z \in \mathrm{L}(\mathcal{Z})$. By substituting according to (6.290), one obtains

$$\begin{aligned} \Phi(Z) &= |\Sigma| \sum_{a,b \in \Gamma} p(a)p(b) \operatorname{vec}(U_a)^* \big(Z \otimes \overline{U_b}\big) \operatorname{vec}(U_a) U_a U_b U_a^* \\ &= |\Sigma| \sum_{a,b \in \Gamma} p(a)p(b) \big\langle U_a U_b U_a^*, Z \big\rangle U_a U_b U_a^*. \end{aligned} \tag{6.315}$$

The natural representation $K(\Phi)$ of the channel Φ is therefore given by

$$\begin{aligned} &|\Sigma| \sum_{a,b \in \Gamma} p(a)p(b) \operatorname{vec}(U_a U_b U_a^*) \operatorname{vec}(U_a U_b U_a^*)^* \\ &= \sum_{a \in \Gamma} p(a) \big(U_a \otimes \overline{U_a}\big) \left(|\Sigma| \sum_{b \in \Gamma} p(b) \operatorname{vec}(U_b) \operatorname{vec}(U_b)^* \right) \big(U_a \otimes \overline{U_a}\big)^* \\ &= \mathbb{1}_{\Sigma} \otimes \mathbb{1}_{\Sigma}, \end{aligned} \tag{6.316}$$

where the last equality has made use of (6.309). It follows that Φ is equal to the identity channel, and therefore statement 2 implies statement 1. □

Theorem 6.53 implies that every mixed-unitary representation of the completely depolarizing channel gives rise to a teleportation protocol, as the following corollary makes precise.

Corollary 6.55. *Let Σ and Γ be alphabets, let $\{U_a : a \in \Gamma\} \subset \mathrm{U}(\mathbb{C}^{\Sigma})$ be a collection of unitary operators, let $p \in \mathcal{P}(\Gamma)$ be a probability vector, and assume that*

$$\Omega(X) = \sum_{a \in \Gamma} p(a) U_a X U_a^* \tag{6.317}$$

for every $X \in \mathrm{L}(\mathbb{C}^{\Sigma})$, where $\Omega \in \mathrm{C}(\mathbb{C}^{\Sigma})$ denotes the completely depolarizing channel with respect to the space $\mathbb{C}^{\Sigma}$. For $\mu : \Gamma \to \mathrm{Pos}(\mathbb{C}^{\Sigma} \otimes \mathbb{C}^{\Sigma})$ defined as

$$\mu(a) = p(a) |\Sigma| \operatorname{vec}(U_a) \operatorname{vec}(U_a)^* \tag{6.318}$$

for each $a \in \Gamma$, one has that μ is a measurement, and moreover

$$X = \frac{1}{|\Sigma|} \sum_{a \in \Gamma} \sum_{b,c \in \Sigma} \langle \mu(a), X \otimes E_{b,c} \rangle \, U_a E_{b,c} U_a^* \tag{6.319}$$

for all $X \in \mathrm{L}(\mathbb{C}^\Sigma)$.

Proof. There is no loss of generality in assuming that $p(a) \neq 0$ for every $a \in \Gamma$, for otherwise one could define an alphabet $\Gamma_0 = \{a \in \Gamma : p(a) \neq 0\}$, verify that the corollary holds in this case, and observe that the statement of the corollary is equivalent when Γ is replaced by Γ_0 in this way.

It is evident that μ is a measurement, as each $\mu(a)$ is positive semidefinite and it holds that

$$\sum_{a \in \Gamma} \mu(a) = \sum_{a \in \Gamma} p(a) |\Sigma| \operatorname{vec}(U_a) \operatorname{vec}(U_a)^* = |\Sigma| J(\Omega) = \mathbb{1}_\Sigma \otimes \mathbb{1}_\Sigma. \tag{6.320}$$

By defining $\Psi_a(X) = U_a X U_a^*$ for every $X \in \mathrm{L}(\mathbb{C}^\Sigma)$ and $a \in \Gamma$, one has that statement 2 of Theorem 6.53 is satisfied. This implies that statement 1 of that theorem holds, which is equivalent to (6.319), and therefore completes the proof. □

Example 6.56. Let $\Sigma = \{0, 1\}$ denote the binary alphabet and let $\Gamma = \Sigma \times \Sigma$. Elements of Γ will be viewed as binary strings of length 2 for convenience. Define $p \in \mathcal{P}(\Gamma)$ as $p(00) = p(01) = p(10) = p(11) = 1/4$ and define unitary operators $U_{00}, U_{01}, U_{10}, U_{11} \in \mathrm{U}(\mathbb{C}^\Sigma)$ as follows:

$$\begin{aligned} U_{00} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad U_{01} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \\ U_{10} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad U_{11} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{6.321}$$

The operators $U_{00}, U_{01}, U_{10}, U_{11}$ coincide with the discrete Weyl operators acting on the space $\mathbb{C}^\Sigma$, and (as explained in Section 4.1.2) provide a mixed-unitary realization of the completely depolarizing channel $\Omega \in \mathrm{C}(\mathbb{C}^\Sigma)$:

$$\frac{1}{4} \sum_{a,b \in \Sigma} U_{ab} X U_{ab}^* = \frac{\operatorname{Tr}(X)}{2} \mathbb{1} \tag{6.322}$$

for every $X \in \mathrm{L}(\mathbb{C}^{\Sigma})$. Consequently, by taking $\mu : \Gamma \to \mathrm{Pos}(\mathbb{C}^{\Sigma} \otimes \mathbb{C}^{\Sigma})$ to be the measurement defined as

$$\begin{aligned}
\mu(00) &= \frac{\operatorname{vec}(U_{00})\operatorname{vec}(U_{00})^*}{2} = u_{00}u_{00}^*, \\
\mu(01) &= \frac{\operatorname{vec}(U_{01})\operatorname{vec}(U_{01})^*}{2} = u_{01}u_{01}^*, \\
\mu(10) &= \frac{\operatorname{vec}(U_{10})\operatorname{vec}(U_{10})^*}{2} = u_{10}u_{10}^*, \\
\mu(11) &= \frac{\operatorname{vec}(U_{11})\operatorname{vec}(U_{11})^*}{2} = u_{11}u_{11}^*,
\end{aligned} \tag{6.323}$$

for

$$\begin{aligned}
u_{00} &= \frac{e_{00} + e_{11}}{\sqrt{2}}, \qquad & u_{01} &= \frac{e_{00} - e_{11}}{\sqrt{2}}, \\
u_{10} &= \frac{e_{01} + e_{10}}{\sqrt{2}}, \qquad & u_{11} &= \frac{e_{01} - e_{10}}{\sqrt{2}},
\end{aligned} \tag{6.324}$$

and setting

$$\Psi_{ab}(X) = U_{ab}XU_{ab}^* \tag{6.325}$$

for each $X \in \mathrm{L}(\mathbb{C}^{\Sigma})$ and $a, b \in \Sigma$, one obtains a teleportation protocol as described in Scenario 6.51. Indeed, the resulting protocol is equivalent to the traditional notion of teleportation in which an ideal single-qubit channel is implemented using a maximally entangled pair of qubits along with two classical bits of communication.

Example 6.57. The previous example may be generalized in the following way. Let $\Sigma = \mathbb{Z}_n$ for any positive integer n, let $\Gamma = \Sigma \times \Sigma$, and let the collection

$$\{U_{ab} : a, b \in \Sigma\} \subset \mathrm{U}(\mathbb{C}^{\Sigma}) \tag{6.326}$$

of unitary operators be in correspondence with the discrete Weyl operators acting on $\mathbb{C}^{\Sigma}$. By taking $\mu : \Gamma \to \mathrm{Pos}(\mathbb{C}^{\Sigma} \otimes \mathbb{C}^{\Sigma})$ to be the measurement defined as

$$\mu(ab) = \frac{\operatorname{vec}(U_{ab})\operatorname{vec}(U_{ab})^*}{n} \tag{6.327}$$

for each $a, b \in \Sigma$, and setting

$$\Psi_{ab}(X) = U_{ab}XU_{ab}^* \tag{6.328}$$

for each $X \in \mathrm{L}(\mathbb{C}^{\Sigma})$, one again obtains a teleportation protocol as described in Scenario 6.51.

In the teleportation protocols described in the previous two examples, the number of distinct classical symbols that must be tramsmitted is equal to the square of the number of classical states in the quantum system that is teleported. This is optimal, as the following corollary states.

Corollary 6.58. *Suppose that Σ and Γ are alphabets, $\mu : \Gamma \to \mathrm{Pos}(\mathbb{C}^{\Sigma} \otimes \mathbb{C}^{\Sigma})$ is a measurement, and $\{\Psi_a : a \in \Gamma\} \subseteq \mathrm{C}(\mathbb{C}^{\Sigma})$ is a collection of channels such that*

$$X = \frac{1}{|\Sigma|} \sum_{a \in \Gamma} \sum_{b,c \in \Sigma} \langle \mu(a), X \otimes E_{b,c} \rangle \Psi_a(E_{b,c}) \tag{6.329}$$

for every $X \in \mathrm{L}(\mathbb{C}^{\Sigma})$. It holds that $|\Gamma| \geq |\Sigma|^2$.

Proof. By Theorem 6.53, it follows that

$$\mu(a) = p(a)|\Sigma| \operatorname{vec}(U_a) \operatorname{vec}(U_a)^* \tag{6.330}$$

for each $a \in \Gamma$, for some choice of a probability vector $p \in \mathcal{P}(\Gamma)$ and a collection of unitary operators $\{U_a : a \in \Gamma\} \subset \mathrm{U}(\mathbb{C}^{\Sigma})$. Each operator $\mu(a)$ has rank at most one, while

$$\sum_{a \in \Gamma} \mu(a) = \mathbb{1}_{\Sigma} \otimes \mathbb{1}_{\Sigma} \tag{6.331}$$

has rank $|\Sigma|^2$. It follows that $|\Gamma| \geq |\Sigma|^2$ as required. □

Dense coding

Along similar lines to the discussion of teleportation above, a scenario in which Alice and Bob aim to implement an ideal classical channel through shared entanglement and quantum communication may be considered.

Scenario 6.59 (Dense coding)**.** Alice holds a register X and Bob holds Y. Both registers have the same classical state set Σ, and the state of the pair (X, Y) is given by the maximally entangled state

$$\tau = \frac{1}{|\Sigma|} \sum_{b,c \in \Sigma} E_{b,c} \otimes E_{b,c}. \tag{6.332}$$

Assume that Alice obtains a classical register Z whose state set is given by some alphabet Γ. She wishes to transmit the classical state $a \in \Gamma$ of Z to Bob by means of a protocol as follows:

1. For $\{\Psi_a : a \in \Gamma\} \subset \mathrm{C}(\mathcal{X})$ being a collection of channels indexed by Γ, Alice performs Ψ_a on X and sends this register to Bob.
2. Bob applies a measurement $\mu : \Gamma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ on the pair (X, Y), and interprets the resulting measurement outcome $b \in \Gamma$ as the outcome of Alice's transmission.

It is not surprising that protocols of this sort exist that function as desired—when Γ is no larger than Σ, the task is trivially accomplished. What is more interesting is that there are protocols of this form that work perfectly in the case that Γ is as large as $\Sigma \times \Sigma$.

The following proposition establishes that a dense coding protocol may be derived from an arbitrary mixed-unitary realization of the completely depolarizing channel, provided the unitary operators are drawn uniformly from a set indexed by $\Sigma \times \Sigma$.

Proposition 6.60. *Let Σ be an alphabet, let $\Gamma = \Sigma \times \Sigma$, let $\mathcal{X} = \mathbb{C}^{\Sigma}$, and let*

$$\tau = \frac{1}{|\Sigma|} \sum_{c,d \in \Sigma} E_{c,d} \otimes E_{c,d}. \tag{6.333}$$

Assume $\{U_a : a \in \Gamma\} \subset \mathrm{U}(\mathcal{X})$ is a collection of unitary operators such that

$$\Omega(X) = \frac{1}{|\Sigma|^2} \sum_{a \in \Gamma} U_a X U_a^* \tag{6.334}$$

for all $X \in \mathrm{L}(\mathcal{X})$, where $\Omega \in \mathrm{C}(\mathcal{X})$ is the completely depolarizing channel with respect to the space $\mathcal{X}$. For $\{\Psi_a : a \in \Gamma\} \subseteq \mathrm{C}(\mathcal{X})$ being a collection of channels defined as

$$\Psi_a(X) = U_a X U_a^* \tag{6.335}$$

for each $a \in \Gamma$ and $X \in \mathrm{L}(\mathcal{X})$, and for $\mu : \Gamma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{X})$ being defined as

$$\mu(a) = \frac{\mathrm{vec}(U_a)\,\mathrm{vec}(U_a)^*}{|\Sigma|} \tag{6.336}$$

for each $a \in \Gamma$, it holds that μ is a measurement and

$$\big\langle \mu(a), (\Psi_b \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(\tau) \big\rangle = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases} \tag{6.337}$$

for all $a, b \in \Gamma$.

Proof. It holds that

$$\sum_{a\in\Gamma}\mu(a) = \frac{1}{|\Sigma|}\sum_{a\in\Gamma}\operatorname{vec}(U_a)\operatorname{vec}(U_a)^* = |\Sigma|J(\Omega) = \mathbb{1}_{\mathcal{X}}\otimes\mathbb{1}_{\mathcal{X}}. \tag{6.338}$$

As each operator $\mu(a)$ is evidently positive semidefinite, it follows that μ is a measurement. For each $a\in\Gamma$, one has

$$\begin{aligned}&\big\langle \mu(a), (\Psi_a\otimes\mathbb{1}_{\mathrm{L}(\mathcal{X})})(\tau)\big\rangle\\ &\quad= \frac{1}{|\Sigma|^2}\big\langle \operatorname{vec}(U_a)\operatorname{vec}(U_a)^*, \operatorname{vec}(U_a)\operatorname{vec}(U_a)^*\big\rangle = 1.\end{aligned} \tag{6.339}$$

Because $(\Psi_b\otimes\mathbb{1}_{\mathrm{L}(\mathcal{X})})(\tau)$ is a density operator for each $b\in\Gamma$, it follows that

$$\big\langle \mu(a), (\Psi_b\otimes\mathbb{1}_{\mathrm{L}(\mathcal{X})})(\tau)\big\rangle = 0 \tag{6.340}$$

for $a\neq b$, which completes the proof. □

Example 6.61. As in Example 6.56, let $\Sigma=\{0,1\}$, let $\Gamma=\Sigma\times\Sigma$, and define unitary operators $U_{00},U_{01},U_{10},U_{11}\in\mathrm{U}(\mathbb{C}^\Sigma)$ as follows:

$$\begin{aligned}U_{00} &= \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \quad U_{01} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\\ U_{10} &= \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \quad U_{11} = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}.\end{aligned} \tag{6.341}$$

As the operators $U_{00},U_{01},U_{10},U_{11}$ provide a mixed-unitary realization of the completely depolarizing channel, by taking $\mu:\Gamma\to\operatorname{Pos}(\mathbb{C}^\Sigma\otimes\mathbb{C}^\Sigma)$ to be the measurement defined as

$$\begin{aligned}\mu(00) &= \frac{\operatorname{vec}(U_{00})\operatorname{vec}(U_{00})^*}{2}, \quad \mu(01) = \frac{\operatorname{vec}(U_{01})\operatorname{vec}(U_{01})^*}{2},\\ \mu(10) &= \frac{\operatorname{vec}(U_{10})\operatorname{vec}(U_{10})^*}{2}, \quad \mu(11) = \frac{\operatorname{vec}(U_{11})\operatorname{vec}(U_{11})^*}{2},\end{aligned} \tag{6.342}$$

and setting $\Psi_{ab}(X) = U_{ab}XU_{ab}^*$ for each $X\in\mathrm{L}(\mathbb{C}^\Sigma)$, as in Example 6.56, one obtains a dense coding protocol as described in Scenario 6.59. The resulting protocol is equivalent to the traditional notion of dense coding in which an ideal two-bit classical channel is implemented using a maximally entangled pair of qubits along with one qubit of communication.

In analogy to the more general type of teleportation protocol described previously, one may consider the capabilities of dense coding protocols for arbitrary choices of an alphabet Γ, as opposed to $\Gamma = \Sigma \times \Sigma$. In particular, suppose Alice's channels are given by the collection $\{\Psi_a : a \in \Gamma\}$, for an arbitrary alphabet Γ, and that the symbol $a \in \Gamma$ Alice wishes to send to Bob is randomly selected according to a probability vector $p \in \mathcal{P}(\Gamma)$. The state of the pair (X, Y) prior to Bob's measurement is described by the ensemble $\eta : \Gamma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{X})$ defined as

$$\eta(a) = \frac{p(a)}{|\Sigma|} \sum_{b,c \in \Sigma} \Psi_a(E_{b,c}) \otimes E_{b,c} \tag{6.343}$$

for all $a \in \Gamma$. The following theorem provides a characterization of when the Holevo information $\chi(\eta)$ of this ensemble attains its maximum possible value, which is $2\log(|\Sigma|)$.

Theorem 6.62. *Let Σ and Γ be alphabets, let $p \in \mathcal{P}(\Gamma)$ be a probability vector such that $p(a) \neq 0$ for all $a \in \Gamma$, and let $\{\Psi_a : a \in \Gamma\} \subseteq \mathrm{C}(\mathbb{C}^\Sigma)$ be a collection of channels. The following two statements are equivalent:*

1. *For the ensemble $\eta : \Gamma \to \mathrm{Pos}(\mathbb{C}^\Sigma \otimes \mathbb{C}^\Sigma)$ defined as*

$$\eta(a) = \frac{p(a)}{|\Sigma|} \sum_{b,c \in \Sigma} \Psi_a(E_{b,c}) \otimes E_{b,c} \tag{6.344}$$

   for all $a \in \Gamma$, one has that $\chi(\eta) = 2\log(|\Sigma|)$.

2. *There exists a collection $\{U_a : a \in \Gamma\} \subset \mathrm{U}(\mathbb{C}^\Sigma)$ of unitary operators with*

$$\Omega(Y) = \sum_{a \in \Gamma} p(a) U_a Y U_a^* \tag{6.345}$$

   for all $Y \in \mathrm{L}(\mathbb{C}^\Sigma)$, where $\Omega \in \mathrm{C}(\mathbb{C}^\Sigma)$ denotes the completely depolarizing channel with respect to the space $\mathbb{C}^\Sigma$, such that $\Psi_a(X) = U_a X U_a^*$ for every choice of $a \in \Gamma$ and $X \in \mathrm{L}(\mathbb{C}^\Sigma)$.

Proof. The Holevo information of the ensemble η defined by (6.344) is

$$\begin{aligned} \chi(\eta) = \mathrm{H}\left(\sum_{a \in \Gamma} \frac{p(a)}{|\Sigma|} \sum_{b,c \in \Sigma} \Psi_a(E_{b,c}) \otimes E_{b,c} \right) \\ - \sum_{a \in \Gamma} p(a)\, \mathrm{H}\left(\frac{1}{|\Sigma|} \sum_{b,c \in \Sigma} \Psi_a(E_{b,c}) \otimes E_{b,c} \right), \end{aligned} \tag{6.346}$$

which may alternatively be written as

$$\chi(\eta) = \mathrm{H}\left(\sum_{a\in\Gamma} p(a)\frac{\mathrm{J}(\Psi_a)}{|\Sigma|}\right) - \sum_{a\in\Gamma} p(a)\,\mathrm{H}\left(\frac{\mathrm{J}(\Psi_a)}{|\Sigma|}\right). \tag{6.347}$$

Under the assumption that $\chi(\eta) = 2\log(|\Sigma|)$, it must hold that

$$\mathrm{H}\left(\sum_{a\in\Gamma} p(a)\frac{\mathrm{J}(\Psi_a)}{|\Sigma|}\right) = 2\log(|\Sigma|) \quad \text{and} \quad \mathrm{H}\left(\frac{\mathrm{J}(\Psi_a)}{|\Sigma|}\right) = 0 \tag{6.348}$$

for each $a \in \Gamma$. The rank of $J(\Psi_a)$ is therefore equal to 1 for each $a \in \Sigma$, and as each Ψ_a is a channel it follows that there must exist a collection of unitary operators $\{U_a : a \in \Gamma\} \subset \mathrm{U}(\mathbb{C}^\Sigma)$ such that $\Psi_a(X) = U_a X U_a^*$ for each $X \in \mathrm{L}(\mathbb{C}^\Sigma)$ and each $a \in \Gamma$. The left equation of (6.348) is equivalent to

$$\sum_{a\in\Gamma} p(a)\frac{\mathrm{J}(\Psi_a)}{|\Sigma|} = \frac{\mathbb{1}\otimes\mathbb{1}}{|\Sigma|^2}, \tag{6.349}$$

which implies

$$\sum_{a\in\Gamma} p(a)\,\mathrm{vec}(U_a)\,\mathrm{vec}(U_a)^* = \frac{\mathbb{1}\otimes\mathbb{1}}{|\Sigma|} = \mathrm{J}(\Omega), \tag{6.350}$$

and therefore

$$\sum_{a\in\Gamma} p(a) U_a Y U_a^* = \Omega(Y) \tag{6.351}$$

for all $Y \in \mathrm{L}(\mathbb{C}^\Sigma)$. Statement 1 therefore implies statement 2.

Under the assumption that statement 2 holds, the Holevo information of η may be calculated directly:

$$\begin{aligned}
\chi(\eta) &= \mathrm{H}\left(\sum_{a\in\Gamma}\frac{p(a)}{|\Sigma|}\sum_{b,c\in\Sigma}\Psi_a(E_{b,c})\otimes E_{b,c}\right) \\
&\qquad - \sum_{a\in\Gamma} p(a)\,\mathrm{H}\left(\frac{1}{|\Sigma|}\sum_{b,c\in\Sigma}\Psi_a(E_{b,c})\otimes E_{b,c}\right) \\
&= \mathrm{H}\left(\frac{\mathbb{1}\otimes\mathbb{1}}{|\Sigma|^2}\right) - \sum_{a\in\Gamma} p(a)\,\mathrm{H}\left(\frac{\mathrm{vec}(U_a)\,\mathrm{vec}(U_a)^*}{|\Sigma|}\right) \\
&= 2\log(|\Sigma|).
\end{aligned} \tag{6.352}$$

Statement 2 therefore implies statement 1, which completes the proof. □

6.3.2 Non-classical correlations

The definition of entanglement, as the absence of separability, is not directly represented by an observable physical phenomenon. There is, however, a fundamental connection between entanglement and the correlations that may exist among the outcomes of measurements performed on two or more separate parts of a physical system. It is helpful to refer to the following scenario when considering this connection.

Scenario 6.63. Two individuals, Alice and Bob, share a compound register (X, Y), with Alice holding X and Bob holding Y. Two events simultaneously occur:

1. Alice receives an input symbol, drawn from a fixed alphabet Σ_{A}, and she must produce an output symbol from a fixed alphabet Γ_{A}.
2. Bob receives an input symbol, drawn from a fixed alphabet Σ_{B}, and he must produce an output symbol from a fixed alphabet Γ_{B}.

Alice and Bob cannot communicate with one another at any point after they have received their input symbols. The output symbols they produce may, in general, be probabilistic, possibly resulting from measurements made on whichever one of the registers X or Y is in the possession of the individual performing the measurement.

The discussion that follows is primarily concerned with the collections of output distributions that may be produced by Alice and Bob, as described in the scenario above, through measurements on a shared entangled state, as compared with the correlations that may result from the initial state of (X, Y) being separable.

Correlation operators

The output distributions produced by Alice and Bob in a particular instance of Scenario 6.63, ranging over all pairs of input symbols, may collectively be described by a single operator

$$C \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{A}} \times \Gamma_{\mathrm{A}}}, \mathbb{R}^{\Sigma_{\mathrm{B}} \times \Gamma_{\mathrm{B}}}\big), \tag{6.353}$$

defined so that $C((a,c),(b,d))$ is the probability that Alice and Bob output $(c,d) \in \Gamma_{\mathrm{A}} \times \Gamma_{\mathrm{B}}$, assuming they are given the input pair $(a,b) \in \Sigma_{\mathrm{A}} \times \Sigma_{\mathrm{B}}$.

Such an operator must satisfy certain constraints. For instance, to carry the interpretation that C represents a collection of probability distributions, each entry must be a nonnegative real number, and it must hold that

$$\sum_{(c,d)\in\Gamma_{\mathrm{A}}\times\Gamma_{\mathrm{B}}} C\big((a,c),(b,d)\big) = 1 \tag{6.354}$$

for every pair $(a,b) \in \Sigma_{\mathrm{A}} \times \Sigma_{\mathrm{B}}$. Additional constraints are imposed by the assumption that Alice and Bob are separated and cannot communicate.

Definition 6.64. Let Σ_{A}, Σ_{B}, Γ_{A}, and Γ_{B} be alphabets, and let

$$C \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{A}}\times\Gamma_{\mathrm{A}}}, \mathbb{R}^{\Sigma_{\mathrm{B}}\times\Gamma_{\mathrm{B}}}\big) \tag{6.355}$$

be an operator.

1. It is said that C is a *deterministic correlation operator* if and only if it takes the form

   $$C = \sum_{(a,b)\in\Sigma_{\mathrm{A}}\times\Sigma_{\mathrm{B}}} E_{a,b} \otimes E_{f(a),g(b)}, \tag{6.356}$$

   or equivalently

   $$C\big((a,c),(b,d)\big) = \begin{cases} 1 & \text{if } c = f(a) \text{ and } d = g(b) \\ 0 & \text{otherwise,} \end{cases} \tag{6.357}$$

   for some choice of functions $f : \Sigma_{\mathrm{A}} \to \Gamma_{\mathrm{A}}$ and $g : \Sigma_{\mathrm{B}} \to \Gamma_{\mathrm{B}}$. It is said that C is a *probabilistic correlation operator* if and only if C is equal to a convex combination of deterministic correlation operators.

2. The operator C is a *quantum correlation operator* if there exist complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, a state $\rho \in \mathrm{D}(\mathcal{X}\otimes\mathcal{Y})$, and two collections of measurements $\{\mu_a : a \in \Sigma_{\mathrm{A}}\}$ and $\{\nu_b : b \in \Sigma_{\mathrm{B}}\}$, taking the form

   $$\mu_a : \Gamma_{\mathrm{A}} \to \mathrm{Pos}(\mathcal{X}) \quad \text{and} \quad \nu_b : \Gamma_{\mathrm{B}} \to \mathrm{Pos}(\mathcal{Y}), \tag{6.358}$$

   such that

   $$C\big((a,c),(b,d)\big) = \big\langle \mu_a(c) \otimes \nu_b(d), \rho \big\rangle \tag{6.359}$$

   for every $a \in \Sigma_{\mathrm{A}}$, $b \in \Sigma_{\mathrm{B}}$, $c \in \Gamma_{\mathrm{A}}$, and $d \in \Gamma_{\mathrm{B}}$.

Example 6.65. Let Σ_{A}, Σ_{B}, Γ_{A}, and Γ_{B} all be equal to the binary alphabet $\Sigma = \{0,1\}$, let $\mathcal{X} = \mathbb{C}^{\Sigma}$ and $\mathcal{Y} = \mathbb{C}^{\Sigma}$, define $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ to be the maximally entangled state

$$\rho = \frac{1}{2} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b}, \tag{6.360}$$

and define measurements $\mu_0, \mu_1 : \Gamma_{\mathrm{A}} \to \mathrm{Pos}(\mathcal{X})$ and $\nu_0, \nu_1 : \Gamma_{\mathrm{B}} \to \mathrm{Pos}(\mathcal{Y})$ as

$$\begin{aligned} \mu_0(0) &= \Pi_0, & \mu_0(1) &= \Pi_{\pi/2}, \\ \mu_1(0) &= \Pi_{\pi/4}, & \mu_1(1) &= \Pi_{3\pi/4}, \\ \nu_0(0) &= \Pi_{\pi/8}, & \nu_0(1) &= \Pi_{5\pi/8}, \\ \nu_1(0) &= \Pi_{7\pi/8}, & \nu_1(1) &= \Pi_{3\pi/8}, \end{aligned} \tag{6.361}$$

for

$$\Pi_\theta = \begin{pmatrix} \cos^2(\theta) & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \sin^2(\theta) \end{pmatrix}. \tag{6.362}$$

Equivalently, these measurement operators are as follows:

$$\begin{aligned} \mu_0(0) &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, & \mu_0(1) &= \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \\ \mu_1(0) &= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, & \mu_1(1) &= \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \\ \nu_0(0) &= \frac{1}{4}\begin{pmatrix} 2+\sqrt{2} & \sqrt{2} \\ \sqrt{2} & 2-\sqrt{2} \end{pmatrix}, & \nu_0(1) &= \frac{1}{4}\begin{pmatrix} 2-\sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 2+\sqrt{2} \end{pmatrix}, \\ \nu_1(0) &= \frac{1}{4}\begin{pmatrix} 2+\sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 2-\sqrt{2} \end{pmatrix}, & \nu_1(1) &= \frac{1}{4}\begin{pmatrix} 2-\sqrt{2} & \sqrt{2} \\ \sqrt{2} & 2+\sqrt{2} \end{pmatrix}. \end{aligned} \tag{6.363}$$

For this choice of ρ, and because each of the measurement operators above have real number entries, it holds that

$$\big\langle \mu_a(c) \otimes \nu_b(d), \rho \big\rangle = \frac{1}{2} \big\langle \mu_a(c), \nu_b(d) \big\rangle \tag{6.364}$$

for each $a \in \Sigma_{\mathrm{A}}$, $b \in \Sigma_{\mathrm{B}}$, $c \in \Gamma_{\mathrm{A}}$, and $d \in \Gamma_{\mathrm{B}}$. A calculation reveals that the

quantum correlation operator defined by (6.359) is given by

$$C = \frac{1}{4}\begin{pmatrix} 2+\sqrt{2} & 2-\sqrt{2} & 2+\sqrt{2} & 2-\sqrt{2} \\ 2-\sqrt{2} & 2+\sqrt{2} & 2-\sqrt{2} & 2+\sqrt{2} \\ 2+\sqrt{2} & 2-\sqrt{2} & 2-\sqrt{2} & 2+\sqrt{2} \\ 2-\sqrt{2} & 2+\sqrt{2} & 2+\sqrt{2} & 2-\sqrt{2} \end{pmatrix}. \tag{6.365}$$

It will be demonstrated shortly that the operator C is not a probabilistic correlation operator.

Example 6.66. Let Σ_A, Σ_B, Γ_A, and Γ_B all be equal to the binary alphabet $\Sigma = \{0,1\}$. There are 16 deterministic correlation operators, which are in correspondence with the 16 possible pairs of functions (f,g) having the form $f : \Sigma_A \to \Gamma_A$ and $g : \Sigma_B \to \Gamma_B$. As matrices, these operators are as follows:

$$\begin{gathered}
\begin{pmatrix} 1&0&1&0\\0&0&0&0\\1&0&1&0\\0&0&0&0 \end{pmatrix}, \begin{pmatrix} 1&0&0&1\\0&0&0&0\\1&0&0&1\\0&0&0&0 \end{pmatrix}, \begin{pmatrix} 0&1&1&0\\0&0&0&0\\0&1&1&0\\0&0&0&0 \end{pmatrix}, \begin{pmatrix} 0&1&0&1\\0&0&0&0\\0&1&0&1\\0&0&0&0 \end{pmatrix}, \\
\begin{pmatrix} 1&0&1&0\\0&0&0&0\\0&0&0&0\\1&0&1&0 \end{pmatrix}, \begin{pmatrix} 1&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&1 \end{pmatrix}, \begin{pmatrix} 0&1&1&0\\0&0&0&0\\0&0&0&0\\0&1&1&0 \end{pmatrix}, \begin{pmatrix} 0&1&0&1\\0&0&0&0\\0&0&0&0\\0&1&0&1 \end{pmatrix}, \\
\begin{pmatrix} 0&0&0&0\\1&0&1&0\\1&0&1&0\\0&0&0&0 \end{pmatrix}, \begin{pmatrix} 0&0&0&0\\1&0&0&1\\1&0&0&1\\0&0&0&0 \end{pmatrix}, \begin{pmatrix} 0&0&0&0\\0&1&1&0\\0&1&1&0\\0&0&0&0 \end{pmatrix}, \begin{pmatrix} 0&0&0&0\\0&1&0&1\\0&1&0&1\\0&0&0&0 \end{pmatrix}, \\
\begin{pmatrix} 0&0&0&0\\1&0&1&0\\0&0&0&0\\1&0&1&0 \end{pmatrix}, \begin{pmatrix} 0&0&0&0\\1&0&0&1\\0&0&0&0\\1&0&0&1 \end{pmatrix}, \begin{pmatrix} 0&0&0&0\\0&1&1&0\\0&0&0&0\\0&1&1&0 \end{pmatrix}, \begin{pmatrix} 0&0&0&0\\0&1&0&1\\0&0&0&0\\0&1&0&1 \end{pmatrix}.
\end{gathered} \tag{6.366}$$

Bell inequalities

By its definition, the set of all probabilistic correlation operators of the form

$$C \in \mathrm{L}\big(\mathbb{R}^{\Sigma_A \times \Gamma_A}, \mathbb{R}^{\Sigma_B \times \Gamma_B}\big) \tag{6.367}$$

is convex. Indeed, this set is given by the convex hull of a finite set, as there are finitely many deterministic correlation operators of the same form. From this fact it follows that the set of all probabilistic correlation operators of the form (6.367) is compact. Therefore, by the separating hyperplane theorem (Theorem 1.11), if an operator

$$D \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{A}} \times \Gamma_{\mathrm{A}}}, \mathbb{R}^{\Sigma_{\mathrm{B}} \times \Gamma_{\mathrm{B}}}\big) \tag{6.368}$$

is not a probabilistic correlation operator, there must exist an operator

$$K \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{A}} \times \Gamma_{\mathrm{A}}}, \mathbb{R}^{\Sigma_{\mathrm{B}} \times \Gamma_{\mathrm{B}}}\big) \tag{6.369}$$

and a real number α such that

$$\langle K, D \rangle > \alpha \quad \text{and} \quad \langle K, C \rangle \leq \alpha \tag{6.370}$$

for all probabilistic correlation operators C of the form (6.367).

For a fixed choice of an operator K and a real number α, the inequality $\langle K, C \rangle \leq \alpha$ is traditionally called a *Bell inequality*, assuming it is satisfied for every probabilistic correlation operator C of the form (6.367). When this is the case, the inequality $\langle K, D \rangle > \alpha$ is called a *Bell inequality violation* if it holds for some choice of a quantum correlation operator D.

The illustration of a Bell inequality violation can provide a convenient way to demonstrate that certain correlation operators are not probabilistic, as the following example illustrates.

Example 6.67 (Clauser–Horn–Shimony–Holt inequality)**.** Let Σ_{A}, Σ_{B}, Γ_{A}, and Γ_{B} all be equal to the binary alphabet $\Sigma = \{0, 1\}$, and define

$$K \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{A}} \times \Gamma_{\mathrm{A}}}, \mathbb{R}^{\Sigma_{\mathrm{B}} \times \Gamma_{\mathrm{B}}}\big) \tag{6.371}$$

as

$$K = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix}. \tag{6.372}$$

For every deterministic correlation operator

$$C \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{A}} \times \Gamma_{\mathrm{A}}}, \mathbb{R}^{\Sigma_{\mathrm{B}} \times \Gamma_{\mathrm{B}}}\big) \tag{6.373}$$

it holds that

$$\langle K, C \rangle \leq 2, \tag{6.374}$$

which may be verified by an inspection of the 16 deterministic correlation operators in Example 6.66. It follows by linearity that the same inequality holds for C being any probabilistic correlation operator. On the other hand, the quantum correlation operator

$$D = \frac{1}{4} \begin{pmatrix} 2+\sqrt{2} & 2-\sqrt{2} & 2+\sqrt{2} & 2-\sqrt{2} \\ 2-\sqrt{2} & 2+\sqrt{2} & 2-\sqrt{2} & 2+\sqrt{2} \\ 2+\sqrt{2} & 2-\sqrt{2} & 2-\sqrt{2} & 2+\sqrt{2} \\ 2-\sqrt{2} & 2+\sqrt{2} & 2+\sqrt{2} & 2-\sqrt{2} \end{pmatrix} \tag{6.375}$$

described in Example 6.65 satisfies

$$\langle K, D \rangle = 2\sqrt{2}. \tag{6.376}$$

This demonstrates that D is not a probabilistic correlation operator.

Correlations among binary-valued measurements

For a given choice of alphabets Σ_A, Σ_B, Γ_A, and Γ_B, and an operator

$$K \in \mathrm{L}\big(\mathbb{R}^{\Sigma_A \times \Gamma_A}, \mathbb{R}^{\Sigma_B \times \Gamma_B}\big), \tag{6.377}$$

it is evidently quite difficult in some cases to determine the supremum value of $\langle K, C \rangle$, optimized over all quantum correlation operators of the form

$$C \in \mathrm{L}\big(\mathbb{R}^{\Sigma_A \times \Gamma_A}, \mathbb{R}^{\Sigma_B \times \Gamma_B}\big). \tag{6.378}$$

There is, however, an interesting class of operators K for which this problem is solvable. This is the class for which the output alphabets Γ_A and Γ_B are both equal to the binary alphabet $\Sigma = \{0, 1\}$, and furthermore the operator K takes the form

$$K = M \otimes \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{6.379}$$

for some choice of an operator

$$M \in \mathrm{L}\big(\mathbb{R}^{\Sigma_B}, \mathbb{R}^{\Sigma_A}\big). \tag{6.380}$$

Operators of the form (6.379) have a simple interpretation when considered in the context of Bell inequalities and violations—they effectively assign the value $M(a,b)$ to the event that Alice and Bob output equal binary-valued answers, and the value $-M(a,b)$ to the event that their outputs differ, for each possible question pair (a,b).

The following theorem, known as Tsirelson's theorem, provides the basis for a solution to the problem under consideration.

Theorem 6.68 (Tsirelson's theorem)**.** *Let $X \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{B}}}, \mathbb{R}^{\Sigma_{\mathrm{A}}}\big)$ be an operator, for alphabets Σ_{A} and Σ_{B}. The following statements are equivalent:*

1. *There exist complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, a state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$, and two collections $\{A_a : a \in \Sigma_{\mathrm{A}}\} \subset \mathrm{Herm}(\mathcal{X})$ and $\{B_b : b \in \Sigma_{\mathrm{B}}\} \subset \mathrm{Herm}(\mathcal{Y})$ of operators satisfying $\|A_a\| \leq 1$, $\|B_b\| \leq 1$, such that*

$$X(a,b) = \big\langle A_a \otimes B_b, \rho \big\rangle \tag{6.381}$$

   for every $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$.

2. *Statement 1 holds under the additional requirement that, for some choice of an alphabet Γ, one has $\mathcal{X} = \mathbb{C}^{\Gamma}$, $\mathcal{Y} = \mathbb{C}^{\Gamma}$, and*

$$\rho = \frac{1}{|\Gamma|} \sum_{c,d \in \Gamma} E_{c,d} \otimes E_{c,d}, \tag{6.382}$$

   and furthermore that the operators in the collections

$$\{A_a : a \in \Sigma_{\mathrm{A}}\} \quad \text{and} \quad \{B_b : b \in \Sigma_{\mathrm{B}}\} \tag{6.383}$$

   are unitary (in addition to being Hermitian).

3. *There exist operators*

$$P \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{A}}}\big) \quad \text{and} \quad Q \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{B}}}\big), \tag{6.384}$$

   with $P(a,a) = 1$ and $Q(b,b) = 1$ for every $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$, such that

$$\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{A}}} \oplus \mathbb{C}^{\Sigma_{\mathrm{B}}}\big). \tag{6.385}$$

4. *There exist two collections $\{u_a : a \in \Sigma_{\mathrm{A}}\}$ and $\{v_b : b \in \Sigma_{\mathrm{B}}\}$ of unit vectors, with elements drawn from the space $\mathbb{R}^{\Sigma_{\mathrm{A}}} \oplus \mathbb{R}^{\Sigma_{\mathrm{B}}}$, such that*

$$X(a,b) = \langle u_a, v_b \rangle \tag{6.386}$$

   for every $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$.

The proof of this theorem will make use of a collection of unitary and Hermitian operators known as *Weyl–Brauer operators*.

Definition 6.69. Let m be a positive integer, let $\Gamma = \{0,1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$. The *Weyl–Brauer operators* $V_0, \ldots, V_{2m} \in \mathrm{L}\big(\mathcal{Z}^{\otimes m}\big)$ of order m are defined as follows:

$$V_0 = \sigma_z^{\otimes m} \tag{6.387}$$

and

$$\begin{aligned} V_{2k-1} &= \sigma_z^{\otimes(k-1)} \otimes \sigma_x \otimes \mathbb{1}^{\otimes(m-k)}, \\ V_{2k} &= \sigma_z^{\otimes(k-1)} \otimes \sigma_y \otimes \mathbb{1}^{\otimes(m-k)}, \end{aligned} \tag{6.388}$$

for $k = 1, \ldots, m$, where $\mathbb{1}$ denotes the identity operator on $\mathcal{Z}$ and σ_x, σ_y, and σ_z are given by the Pauli operators. In matrix form, these operators are as follows:

$$\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6.389}$$

Example 6.70. In the case $m = 3$, the Weyl–Brauer operators $V_0, \ldots, V_6$ are

$$\begin{aligned} V_0 &= \sigma_z \otimes \sigma_z \otimes \sigma_z \\ V_1 &= \sigma_x \otimes \mathbb{1} \otimes \mathbb{1} \\ V_2 &= \sigma_y \otimes \mathbb{1} \otimes \mathbb{1} \\ V_3 &= \sigma_z \otimes \sigma_x \otimes \mathbb{1} \\ V_4 &= \sigma_z \otimes \sigma_y \otimes \mathbb{1} \\ V_5 &= \sigma_z \otimes \sigma_z \otimes \sigma_x \\ V_6 &= \sigma_z \otimes \sigma_z \otimes \sigma_y. \end{aligned} \tag{6.390}$$

A proposition summarizing the properties of the Weyl–Brauer operators that are relevant to the proof of Tsirelson's theorem follows.

Proposition 6.71. *Let m be a positive integer, let $V_0, \ldots, V_{2m}$ denote the Weyl–Brauer operators of order m, and let $(\alpha_0, \ldots, \alpha_{2m}), (\beta_0, \ldots, \beta_{2m}) \in \mathbb{R}^{2m+1}$ be vectors of real numbers. It holds that*

$$\left(\sum_{k=0}^{2m} \alpha_k V_k\right)^2 = \left(\sum_{k=0}^{2m} \alpha_k^2\right) \mathbb{1}^{\otimes m} \tag{6.391}$$

and

$$\frac{1}{2^m}\left\langle \sum_{j=0}^{2m} \alpha_j V_j, \sum_{k=0}^{2m} \beta_k V_k \right\rangle = \sum_{k=0}^{2m} \alpha_k \beta_k. \tag{6.392}$$

Proof. The Pauli operators anti-commute in pairs:

$$\sigma_x \sigma_y = -\sigma_y \sigma_x, \quad \sigma_x \sigma_z = -\sigma_z \sigma_x, \quad \text{and} \quad \sigma_y \sigma_z = -\sigma_z \sigma_y. \tag{6.393}$$

By an inspection of the definition of the Weyl–Brauer operators, it follows that $V_0, \ldots, V_{2m}$ also anti-commute in pairs:

$$V_j V_k = -V_k V_j \tag{6.394}$$

for distinct choices of $j, k \in \{0, \ldots, 2m\}$. Moreover, each V_k is both unitary and Hermitian, and therefore $V_k^2 = \mathbb{1}^{\otimes m}$. It follows that

$$\begin{aligned} \left(\sum_{k=0}^{2m} \alpha_k V_k \right)^2 &= \sum_{k=0}^{2m} \alpha_k^2 V_k^2 + \sum_{0 \leq j < k \leq 2m} \alpha_j \alpha_k \big(V_j V_k + V_k V_j \big) \\ &= \left(\sum_{k=0}^{2m} \alpha_k^2 \right) \mathbb{1}^{\otimes m}. \end{aligned} \tag{6.395}$$

Moreover,

$$\langle V_j, V_k \rangle = \begin{cases} 2^m & \text{if } j = k \\ 0 & \text{if } j \neq k, \end{cases} \tag{6.396}$$

and therefore

$$\frac{1}{2^m}\left\langle \sum_{j=0}^{2m} \alpha_j V_j, \sum_{k=0}^{2m} \beta_k V_k \right\rangle = \frac{1}{2^m} \sum_{j=0}^{2m} \sum_{k=0}^{2m} \alpha_j \beta_k \langle V_j, V_k \rangle = \sum_{k=0}^{2m} \alpha_k \beta_k, \tag{6.397}$$

as required. □

Proof of Theorem 6.68. The implications to be proved among the statements, which suffice to prove the theorem, may be summarized as follows:

$$(2) \Rightarrow (1) \Rightarrow (3) \Rightarrow (4) \Rightarrow (2). \tag{6.398}$$

One has that statement 2 trivially implies statement 1.

Assume statement 1 holds, define an operator

$$K = \sum_{a \in \Sigma_{\mathrm{A}}} e_a \operatorname{vec}\big((A_a \otimes \mathbb{1}) \sqrt{\rho} \big)^* + \sum_{b \in \Sigma_{\mathrm{B}}} e_b \operatorname{vec}\big((\mathbb{1} \otimes B_b) \sqrt{\rho} \big)^*, \tag{6.399}$$

and consider the operator $KK^* \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{A}} \sqcup \Sigma_{\mathrm{B}}}\big)$, which may be written in a block form as

$$KK^* = \begin{pmatrix} P & Y \\ Y^* & Q \end{pmatrix} \tag{6.400}$$

for $P \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{A}}}\big)$, $Q \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{B}}}\big)$, and $Y \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{B}}}, \mathbb{C}^{\Sigma_{\mathrm{A}}}\big)$. It holds that

$$Y(a,b) = \big\langle (A_a \otimes \mathbb{1})\sqrt{\rho}, (\mathbb{1} \otimes B_b)\sqrt{\rho} \big\rangle = \big\langle A_a \otimes B_b, \rho \big\rangle = X(a,b) \tag{6.401}$$

for every $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$, and therefore $Y = X$. Moreover, for each $a \in \Sigma_{\mathrm{A}}$ one has

$$P(a,a) = \big\langle (A_a \otimes \mathbb{1})\sqrt{\rho}, (A_a \otimes \mathbb{1})\sqrt{\rho} \big\rangle = \big\langle A_a^2 \otimes \mathbb{1}, \rho \big\rangle, \tag{6.402}$$

which is necessarily a nonnegative real number in the interval $[0,1]$; and through a similar calculation, one finds that $Q(b,b)$ is also a nonnegative integer in the interval $[0,1]$ for each $b \in \Sigma_{\mathrm{B}}$. A nonnegative real number may be added to each diagonal entry of this operator to yield another positive semidefinite operator, so one has that statement 3 holds. It has therefore been proved that statement 1 implies statement 3.

Next, assume statement 3 holds, and observe that

$$\frac{1}{2}\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} + \frac{1}{2}\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} \frac{P+\overline{P}}{2} & X \\ X^* & \frac{Q+\overline{Q}}{2} \end{pmatrix} \tag{6.403}$$

is a positive semidefinite operator having real number entries, and all of its diagonal entries are equal to 1. Define

$$u_a = \begin{pmatrix} \frac{P+\overline{P}}{2} & X \\ X^* & \frac{Q+\overline{Q}}{2} \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} e_a \\ 0 \end{pmatrix} \quad \text{and} \quad v_b = \begin{pmatrix} \frac{P+\overline{P}}{2} & X \\ X^* & \frac{Q+\overline{Q}}{2} \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} 0 \\ e_b \end{pmatrix} \tag{6.404}$$

for each $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$. As the square root of a positive semidefinite operator having real number entries also has real number entries, one has that u_a and v_b are unit vectors with real number entries, and moreover it holds that

$$\langle u_a, v_b \rangle = X(a,b) \tag{6.405}$$

for all $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$. It has therefore been proved that statement 3 implies statement 4.

Finally, assume statement 4 holds. Let

$$m = \left\lceil \frac{|\Sigma_{\mathrm{A}}| + |\Sigma_{\mathrm{B}}| - 1}{2} \right\rceil, \tag{6.406}$$

so that $2m + 1 \geq |\Sigma_{\mathrm{A}}| + |\Sigma_{\mathrm{B}}|$, and let $f : \Sigma_{\mathrm{A}} \sqcup \Sigma_{\mathrm{B}} \to \{0, \ldots, 2m\}$ be a fixed but otherwise arbitrarily chosen one-to-one function. Let $\Gamma = \{0, 1\}$, let $\mathcal{Z} = \mathbb{C}^{\Gamma}$, and define

$$A_a = \sum_{c \in \Sigma_{\mathrm{A}} \sqcup \Sigma_{\mathrm{B}}} u_a(c) V_{f(c)} \quad \text{and} \quad B_b = \sum_{c \in \Sigma_{\mathrm{A}} \sqcup \Sigma_{\mathrm{B}}} v_b(c) V_{f(c)}^{\mathrm{T}} \tag{6.407}$$

for each $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$, for $V_0, \ldots, V_{2m}$ being the Weyl–Brauer operators of order m, regarded as operators acting on $\mathcal{Z}^{\otimes m}$. As the vectors

$$\{u_a : a \in \Sigma_{\mathrm{A}}\} \quad \text{and} \quad \{v_b : b \in \Sigma_{\mathrm{B}}\} \tag{6.408}$$

are unit vectors having real number entries, it follows from Proposition 6.71 that the operators

$$\{A_a : a \in \Sigma_{\mathrm{A}}\} \quad \text{and} \quad \{B_b : b \in \Sigma_{\mathrm{B}}\} \tag{6.409}$$

are unitary, and it is evident that they are Hermitian as well. Define

$$\tau = \frac{1}{2^m} \operatorname{vec}\bigl(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\bigr) \operatorname{vec}\bigl(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\bigr)^{*}. \tag{6.410}$$

For each choice of $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$ it holds that

$$\begin{aligned} &\bigl\langle A_a \otimes B_b, \tau \bigr\rangle = \frac{1}{2^m} \operatorname{Tr}\bigl(A_a B_b^{\mathrm{T}}\bigr) \\ &\frac{1}{2^m} \sum_{c,d \in \Sigma_{\mathrm{A}} \sqcup \Sigma_{\mathrm{B}}} \bigl\langle u_a(c) V_{f(c)}, v_b(d) V_{f(d)} \bigr\rangle = \langle u_a, v_b \rangle, \end{aligned} \tag{6.411}$$

again by Proposition 6.71. This is equivalent to statement 2 (taking Γ^m in place of Γ). It has therefore been proved that statement 4 implies statement 2, which completes the proof. □

As a consequence of Tsirelson's theorem (Theorem 6.68), there exists a semidefinite program for the supremum value of the inner product $\langle K, C \rangle$, for K taking the form (6.379) and for C ranging over all quantum correlation operators of the form

$$C \in \mathrm{L}\bigl(\mathbb{R}^{\Sigma_{\mathrm{A}} \times \Gamma_{\mathrm{A}}}, \mathbb{R}^{\Sigma_{\mathrm{B}} \times \Gamma_{\mathrm{B}}}\bigr), \tag{6.412}$$

for Σ_{A} and Σ_{B} being arbitrary alphabets and Γ_{A} and Γ_{B} both being equal to the binary alphabet $\Gamma = \{0,1\}$.

To understand why this is so, consider an arbitrary quantum correlation operator C, which must be given by

$$C\big((a,c),(b,d)\big) = \big\langle \mu_a(c) \otimes \nu_b(d), \rho \big\rangle \tag{6.413}$$

for every $a \in \Sigma_{\mathrm{A}}$, $b \in \Sigma_{\mathrm{B}}$, and $c, d \in \Gamma$, for some choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, a state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$, and two collections of measurements $\{\mu_a : a \in \Sigma_{\mathrm{A}}\}$ and $\{\nu_b : b \in \Sigma_{\mathrm{B}}\}$ whose elements take the form

$$\mu_a : \Gamma \to \mathrm{Pos}(\mathcal{X}) \quad \text{and} \quad \nu_b : \Gamma \to \mathrm{Pos}(\mathcal{Y}). \tag{6.414}$$

For an operator K of the form (6.379) for some choice of $M \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{B}}}, \mathbb{R}^{\Sigma_{\mathrm{A}}}\big)$, one has that the value of the inner product $\langle K, C \rangle$ is given by

$$\sum_{(a,b) \in \Sigma_{\mathrm{A}} \times \Sigma_{\mathrm{B}}} M(a,b) \big\langle (\mu_a(0) - \mu_a(1)) \otimes (\nu_b(0) - \nu_b(1)), \rho \big\rangle. \tag{6.415}$$

Now, an operator H, acting on an arbitrary complex Euclidean space, may be written as

$$H = \mu(0) - \mu(1) \tag{6.416}$$

for some binary-valued measurement μ if and only if H is Hermitian and satisfies $\|H\| \leq 1$. Thus, an optimization of the expression (6.415) over all choices of the measurements $\{\mu_a : a \in \Sigma_{\mathrm{A}}\}$ and $\{\nu_b : b \in \Sigma_{\mathrm{B}}\}$ is equivalent to an optimization of the expression

$$\sum_{(a,b) \in \Sigma_{\mathrm{A}} \times \Sigma_{\mathrm{B}}} M(a,b) \big\langle A_a \otimes B_b, \rho \big\rangle \tag{6.417}$$

over all collections

$$\{A_a : a \in \Sigma_{\mathrm{A}}\} \subset \mathrm{Herm}(\mathcal{X}) \quad \text{and} \quad \{B_b : b \in \Sigma_{\mathrm{B}}\} \subset \mathrm{Herm}(\mathcal{Y}) \tag{6.418}$$

of Hermitian operators satisfying $\|A_a\| \leq 1$ and $\|B_b\| \leq 1$, for every $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$, respectively.

By optimizing over all complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$ and density operators $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$, one finds (by Theorem 6.68) that the supremum value of $\langle K, C \rangle$ over all quantum correlation operators C is equal to the

supremum value of the inner product $\langle M, X \rangle$ over all choices of operators $X \in \mathrm{L}\big(\mathbb{R}^{\Sigma_{\mathrm{B}}}, \mathbb{R}^{\Sigma_{\mathrm{A}}}\big)$ for which it holds that

$$\begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{A}}} \oplus \mathbb{C}^{\Sigma_{\mathrm{B}}}\big), \tag{6.419}$$

for $P \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{A}}}\big)$ and $Q \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{B}}}\big)$ satisfying $P(a,a) = 1$ and $Q(b,b) = 1$ for every $a \in \Sigma_{\mathrm{A}}$ and $b \in \Sigma_{\mathrm{B}}$. Such an optimization corresponds directly to the following primal problem of a semidefinite program:

$$\begin{aligned} &\underline{\text{Primal problem}} \\ \text{maximize:} \quad & \frac{1}{2}\big\langle M, X \big\rangle + \frac{1}{2}\big\langle M^*, X^* \big\rangle \\ \text{subject to:} \quad & \begin{pmatrix} P & X \\ X^* & Q \end{pmatrix} \geq 0, \\ & \Delta(P) = \mathbb{1}, \\ & \Delta(Q) = \mathbb{1}, \\ & P \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{A}}}\big), \\ & Q \in \mathrm{Pos}\big(\mathbb{C}^{\Sigma_{\mathrm{B}}}\big), \\ & X \in \mathrm{L}\big(\mathbb{C}^{\Sigma_{\mathrm{B}}}, \mathbb{C}^{\Sigma_{\mathrm{A}}}\big). \end{aligned}$$

In this problem, Δ refers to the completely dephasing channel, defined with respect to either $\mathbb{C}^{\Sigma_{\mathrm{A}}}$ or $\mathbb{C}^{\Sigma_{\mathrm{B}}}$, and $\mathbb{1}$ denotes the identity operator on either of these spaces, as the context dictates without ambiguity.

The dual problem of this semidefinite program is as follows:

$$\begin{aligned} &\underline{\text{Dual problem}} \\ \text{minimize:} \quad & \frac{1}{2}\,\mathrm{Tr}(Y) + \frac{1}{2}\,\mathrm{Tr}(Z) \\ \text{subject to:} \quad & \begin{pmatrix} \Delta(Y) & -M \\ -M^* & \Delta(Z) \end{pmatrix} \geq 0, \\ & Y \in \mathrm{Herm}\big(\mathbb{C}^{\Sigma_{\mathrm{A}}}\big), \\ & Z \in \mathrm{Herm}\big(\mathbb{C}^{\Sigma_{\mathrm{B}}}\big). \end{aligned}$$

It follows from Slater's theorem (Theorem 1.13) that strong duality holds for this semidefinite program—strict feasibility holds for both the primal and dual problems.

Example 6.72 (Tsirelson's bound)**.** Consider the operator

$$K = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix} = M \otimes \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{6.420}$$

for

$$M = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{6.421}$$

which was examined in Example 6.67. One has $\|M\| = \sqrt{2}$, so that

$$\begin{pmatrix} \sqrt{2}\mathbb{1} & -M \\ -M^* & \sqrt{2}\mathbb{1} \end{pmatrix} \geq 0. \tag{6.422}$$

By taking $Y = \sqrt{2}\mathbb{1}$ and $Z = \sqrt{2}\mathbb{1}$ in the dual problem above, a feasible dual solution achieving the objective value $2\sqrt{2}$ is obtained. Therefore,

$$\langle K, C \rangle \leq 2\sqrt{2} \tag{6.423}$$

for every quantum correlation operator C. The Bell inequality violation exhibited in Example 6.67 is therefore optimal for this choice of K.

6.4 Exercises

6.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Prove that the following three statements are equivalent:

1. For every complex Euclidean space $\mathcal{Z}$ and every state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Z})$, it holds that
$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)(\rho) \in \mathrm{SepD}(\mathcal{Y} : \mathcal{Z}). \tag{6.424}$$
2. $J(\Phi) \in \mathrm{Sep}(\mathcal{Y} : \mathcal{X})$.
3. There exists an alphabet Σ, a measurement $\mu : \Sigma \to \mathrm{Pos}(\mathcal{X})$, and a collection of states $\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y})$ such that
$$\Phi(X) = \sum_{a \in \Sigma} \langle \mu(a), X \rangle \sigma_a \tag{6.425}$$
for all $X \in \mathrm{L}(\mathcal{X})$.

Channels for which these statements hold are called *entanglement-breaking* channels.

6.2. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, both being of dimension n, let $\{U_1, \ldots, U_m\} \in \mathrm{U}(\mathcal{Y}, \mathcal{X})$ be a set of pairwise orthogonal isometries, and let $u_k \in \mathcal{X} \otimes \mathcal{Y}$ be the vector defined as

$$u_k = \frac{1}{\sqrt{n}} \operatorname{vec}(U_k) \tag{6.426}$$

for each $k \in \{1, \ldots, m\}$. Let $\mu : \{1, \ldots, m\} \to \operatorname{Pos}(\mathcal{X} \otimes \mathcal{Y})$ be a measurement such that $\mu(k) \in \operatorname{PPT}(\mathcal{X} : \mathcal{Y})$ for every $k \in \{1, \ldots, m\}$. Prove that

$$\sum_{k=1}^{m} \langle \mu(k), u_k u_k^* \rangle \leq n. \tag{6.427}$$

(Observe that a correct solution to this exercise generalizes Theorem 6.33.)

6.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces having dimension at least 2. Prove that there exist entanglement-breaking channels $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, as defined in Exercise 6.1, such that

$$\bigl|\!\bigl|\!\bigl| \Phi_0 - \Phi_1 \bigr|\!\bigr|\!\bigr|_1 > \bigl\| \Phi_0(\rho) - \Phi_1(\rho) \bigr\|_1 \tag{6.428}$$

for every $\rho \in \mathrm{D}(\mathcal{X})$. Such channels have the seemingly strange property that they destroy entanglement, and yet evaluating them on an entangled state helps to discriminate between them.

6.4. Let X and Y be registers and let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be a state of the pair (X, Y). With respect to ρ, one defines the *entanglement of formation* between the pair (X, Y) as

$$\mathrm{E}_{\mathrm{F}}(\mathsf{X} : \mathsf{Y}) = \inf \left\{ \sum_{a \in \Sigma} p(a)\, \mathrm{H}\bigl(\operatorname{Tr}_{\mathcal{Y}}(u_a u_a^*)\bigr) \;:\; \sum_{a \in \Sigma} p(a) u_a u_a^* = \rho \right\}, \tag{6.429}$$

where the infimum is over all choices of an alphabet Σ, a probability vector $p \in \mathcal{P}(\Sigma)$, and a collection of unit vectors $\{u_a : a \in \Sigma\} \subset \mathcal{X} \otimes \mathcal{Y}$ for which it holds that

$$\sum_{a \in \Sigma} p(a) u_a u_a^* = \rho. \tag{6.430}$$

(a) Prove that the infimum in (6.429) is achieved for some choice of Σ, p, and $\{u_a : a \in \Sigma\}$ for which $|\Sigma| \leq \dim(\mathcal{X} \otimes \mathcal{Y})^2$.

(b) Prove that $\mathrm{E}_{\mathrm{D}}(\mathsf{X}:\mathsf{Y}) \leq \mathrm{E}_{\mathrm{F}}(\mathsf{X}:\mathsf{Y}) \leq \mathrm{E}_{\mathrm{C}}(\mathsf{X}:\mathsf{Y})$.

(c) Suppose that Z and W are registers and $\Phi \in \mathrm{LOCC}(\mathcal{X},\mathcal{Z}:\mathcal{Y},\mathcal{W})$ is an LOCC channel. Prove that

$$\mathrm{E}_{\mathrm{F}}(\mathsf{Z}:\mathsf{W})_{\sigma} \leq \mathrm{E}_{\mathrm{F}}(\mathsf{X}:\mathsf{Y})_{\rho} \tag{6.431}$$

where $\sigma = \Phi(\rho)$ and $\mathrm{E}_{\mathrm{F}}(\mathsf{X}:\mathsf{Y})_{\rho}$ and $\mathrm{E}_{\mathrm{F}}(\mathsf{Z}:\mathsf{W})_{\sigma}$ denote the entanglement of formation of the pairs (X,Y) and (Z,W) with respect to the states ρ and σ, respectively.

(d) Prove a more general statement than the one required of a solution to part (c), holding not only for all LOCC channels, but for all *separable* channels of the form $\Phi \in \mathrm{SepC}(\mathcal{X},\mathcal{Z}:\mathcal{Y},\mathcal{W})$.

6.5. Let Σ be an alphabet, let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces of the form $\mathcal{X} = \mathbb{C}^{\Sigma}$ and $\mathcal{Y} = \mathbb{C}^{\Sigma}$, let $n = |\Sigma|$, and consider the projection operators Δ_0, Δ_1, Π_0, and Π_1 defined in Example 6.11. The states

$$\rho_0 = \frac{\Pi_0}{\binom{n+1}{2}} \quad \text{and} \quad \rho_1 = \frac{\Pi_1}{\binom{n}{2}} \tag{6.432}$$

are therefore Werner states, while

$$\sigma_0 = \Delta_0 \quad \text{and} \quad \sigma_1 = \frac{\Delta_1}{n^2 - 1} \tag{6.433}$$

are isotropic states.

(a) Prove that if $\mu : \{0,1\} \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ is a measurement satisfying $\mu(0), \mu(1) \in \mathrm{PPT}(\mathcal{X}:\mathcal{Y})$, then

$$\frac{1}{2}\langle \mu(0), \rho_0 \rangle + \frac{1}{2}\langle \mu(1), \rho_1 \rangle \leq \frac{1}{2} + \frac{1}{n+1}. \tag{6.434}$$

Prove that there exists an LOCC measurement μ for which (6.434) holds with equality.

(b) Prove that if $\nu : \{0,1\} \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ is a measurement satisfying $\nu(0), \nu(1) \in \mathrm{PPT}(\mathcal{X}:\mathcal{Y})$, then

$$\frac{1}{2}\langle \nu(0), \sigma_0 \rangle + \frac{1}{2}\langle \nu(1), \sigma_1 \rangle \leq 1 - \frac{1}{2n+2}. \tag{6.435}$$

Prove that there exists an LOCC measurement ν for which (6.435) holds with equality.

6.6. Let N and m be positive integers, and assume that there exist unitary and Hermitian operators $U_0, \ldots, U_{2m} \in \mathrm{L}(\mathbb{C}^N)$ that anti-commute in pairs: $U_j U_k = -U_k U_j$ for distinct choices of $j, k \in \{0, \ldots, 2m\}$. Prove that the collection

$$\left\{ U_0^{a_0} \cdots U_{2m}^{a_{2m}} \,:\, a_0, \ldots, a_{2m} \in \{0,1\},\ a_0 + \cdots + a_{2m} \text{ is even} \right\} \tag{6.436}$$

is an orthogonal collection, and conclude that $N \geq 2^m$.

Observe that a correct solution to this exercise implies that the Weyl–Brauer operators have the minimum possible dimension required to possess the properties mentioned above.

6.5 Bibliographic remarks

The phenomenon of entanglement was first recognized in a 1935 paper of Einstein, Podolsky, and Rosen [68], although it was not formally defined or called entanglement therein. Einstein, Podolsky, and Rosen's work inspired Schrödinger to investigate the phenomenon of entanglement, and to give it its name; he published a three-part paper in German [180, 181, 182], as well as two related English-language papers [183, 184] discussing entanglement and other issues, as they pertained to the nature of quantum physics at that time. (An English translation of Schrödinger's three-part paper in German was published later [204].) The identification of entanglement with a lack of separability is due to Werner [225], who used the terms *classically correlated* and *EPR correlated* rather than *separable* and *entangled*.

The equivalence of the first two statements in Theorem 6.10 was proved by M. Horodecki, P. Horodecki, and R. Horodecki [116], and Proposition 6.7 was proved by P. Horodecki [119]. Several elementary analytic facts about the set of separable states that have been discussed in Section 6.1.1 were also observed in the papers proving these facts. The equivalence of the third statement in Theorem 6.10 to the first two was proved a few years later by P. Horodecki [120]. In general, it is likely to be a computationally difficult task to test a bipartite density operator for separability, as suggested by the hardness result proved by Gurvits [82].

The fact that any operator sufficiently close to the identity operator in a bipartite tensor product space is separable was first proved by Życzkowski, P. Horodecki, Sanpera, and Lewenstein [237]. Theorem 6.14 is due to Gurvits and Barnum [83].

The local operations and classical communication paradigm, also called the *distant labs* paradigm, arose naturally in quantum information theory as various quantum information processing tasks were considered. Among the first researchers to consider this paradigm were Peres and Wootters [171], who compared the capabilities of LOCC measurements to general measurements in a setting in which information is encoded into bipartite product states. The teleportation procedure of Bennett, Brassard, Crepéau, Josza, Peres, and Wootters [31]—certainly one of the most important LOCC procedures, both historically and theoretically speaking—followed shortly after.

There are natural extensions of the definition of LOCC channels that have not been discussed in this chapter. In particular, the definition of LOCC channels in the present chapter requires an LOCC channel to be a finite composition of one-way LOCC channels, corresponding to a fixed number of classical message transmissions between two individuals implementing the channel, but one may also consider channels implemented by a potentially unbounded number of message transmissions. It is known that the set of LOCC channels, as they have been defined in this chapter, is generally not closed for a fixed choice of spaces; this was proved (for bipartite channels) by Chitambar, Leung, Mančinska, Ozols, and Winter [49]. The definition of LOCC channels presented in this chapter is based on one of the definitions considered by these authors.

The class of separable channels was identified by Vedral, Plenio, Rippin, and Knight [212], although they did not raise the possibility (first suggested by Rains [174]) that some separable channels might not be LOCC channels. The existence of separable measurements that are not LOCC measurements (and, in fact, not even approached by a sequence of LOCC measurements in the limit) was proved by Bennett, DiVincenzo, Fuchs, Mor, Rains, Shor, Smolin, and Wootters [33]. Childs, Leung, Mančinska, and Ozols [45] give a simplified proof of this fact, along with some generalizations of it.

Bennett, Bernstein, Popescu, and Schumacher [30] defined the distillable entanglement and entanglement cost, and proved Theorem 6.42 through the design and analysis of LOCC channels for entanglement distillation and its reverse for pure states. (Bennett, Bernstein, Popescu, and Schumacher used the term *entanglement of formation* rather than entanglement cost—but that terminology has since come to refer to a related but different measure of entanglement.)

Entanglement distillation for general quantum states was considered by Bennett, Brassard, Popescu, Schumacher, Smolin, and Wootters [32] and Bennett, DiVincenzo, Smolin, and Wootters [36] around the same time. It is known that the entanglement cost of every bipartite entangled state is nonzero [233].

The entanglement rank was first defined by Terhal and P. Horodecki [200], who referred to it as the *Schmidt number* of a density operator (as it generalizes the number of nonzero terms in a Schmidt decomposition of the vector representation of a given pure state). They also proved that the entanglement rank of a state cannot increase under the action of an LOCC channel, based on related observations by Lo and Popescu [151] regarding pure states, and that it is generally not multiplicative with respect to tensor products.

Theorem 6.33 was proved by Nathanson [160], and Theorem 6.35 was proved by Walgate, Short, Hardy, and Vedral [218].

The equivalence of statements 1, 2, and 3 in Theorem 6.37, as well as statement 4 for LOCC channels rather than separable channels, was proved by Nielsen [163]. Nielsen's proof made use of the fact that every bipartite pure state transformation induced by an LOCC channel is also induced by a one-way LOCC channel, which was proved earlier by Lo and Popescu [151]. The equivalent of statement 4 of Nielsen's theorem with the first three was proved by Gheorghiu and Griffiths [77]. The proof of Theorem 6.42 concerning entanglement distillation and cost for pure states also appears in the same paper of Nielsen.

Peres [170] proposed the computationally efficient partial transpose test for separability of bipartite density operators; he observed that separable states are necessarily PPT, and that interesting families of entangled states were revealed to be entangled through this test. By the Horodecki criterion (Theorem 6.10) proved shortly after, it follows that the partial transpose test correctly identifies all entangled state in a tensor product of two complex Euclidean spaces, both of dimension 2 or one of dimension 2 and one of dimension 3, based on work of Størmer [198] and Woronowicz [232], but that entangled PPT states in higher dimensions must exist [116]. The first explicit examples of entangled PPT states were given by P. Horodecki [119]; the unextendable product set construction of such states is due to Bennett, DiVincenzo, Mor, Shor, Smolin, and Terhal [34], who introduced the notion of an unextendable product set as well as the specific example given in this

chapter. Proposition 6.49 and Theorem 6.50 were proved by M. Horodecki, P. Horodecki, and R. Horodecki [117].

The teleportation procedure described in Example 6.56 is, as mentioned above, due to Bennett, Brassard, Crepéau, Josza, Peres, and Wootters [31]. The dense coding procedure described in Example 6.61 is due to Bennett and Wiesner [40]. Various generalizations of these procedures have been discovered—the general presentation of teleportation and dense coding in this chapter is based on work of Werner [227].

The fact that entangled states may induce non-classical correlations was discovered by Bell in a highly influential 1964 paper [27]. The Bell inequality described in Example 6.67 is due to Clauser, Horn, Shimony, and Holt [52]. Some entangled states fail to induce non-classical correlations—this was demonstrated for the special case in which only projective measurements are made on the two parts of a bipartite state by Werner [225], and for general measurements by Barrett [23]. The entangled states constructed by Werner that have this property are among those described in Example 6.11. Theorem 6.68 is due to Tsirelson [205].

This chapter has presented just a small part of an extensive body of work on entanglement. Readers interested in learning more about this topic are referred to the survey of R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki [121].

Chapter 7

Permutation invariance and unitarily invariant measures

This chapter introduces two notions—*permutation invariance* and *unitarily invariant measures*—having interesting applications in quantum information theory. A state of a collection of identical registers is said to be permutation invariant if it is unchanged under arbitrary permutations of the contents of the registers. Unitarily invariant measures are Borel measures, defined for sets of vectors or operators, that are unchanged by the action of all unitary operators acting on the underlying space. The two notions are distinct but nevertheless linked, with the interplay between them offering a useful tool for performing calculations in both settings.

7.1 Permutation-invariant vectors and operators

This section of the chapter discusses properties of permutation invariant states of collections of identical registers. Somewhat more generally, one may consider permutation invariant positive semidefinite operators, as well as permutation invariant vectors having arbitrary norm.

It is to be assumed for the entirety of the section that an alphabet Σ and a positive integer $n \geq 2$ have been fixed, and that $\mathsf{X}_1, \ldots, \mathsf{X}_n$ is a sequence of registers sharing the same classical state set Σ. Algebraic properties of states of the compound register $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ relating to permutations and symmetries among the individual registers will be a primary focus.

7.1.1 The subspace of permutation-invariant vectors

Within the tensor product space

$$\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \tag{7.1}$$

which is equal to the n-fold tensor product of the space $\mathbb{C}^\Sigma$ with itself, some vectors are unchanged under all possible permutations among the tensor factors $\mathcal{X}_1, \ldots, \mathcal{X}_n$. The set of all such vectors forms a subspace, known as the *symmetric subspace*. A more formal description of this subspace will be given shortly, following a short discussion of those operators that represent permutations among the tensor factors of the space (7.1).

Permutations of tensor factors

For each choice of $k \in \{1, \ldots, n\}$, the complex Euclidean space $\mathcal{X}_k$ associated with the register X_k, as described above, is equal to $\mathbb{C}^\Sigma$, and for this reason an arbitrary vector $u \in \mathbb{C}^\Sigma$ may be regarded as an element of any one of the spaces $\mathcal{X}_1, \ldots, \mathcal{X}_n$. With this fact in mind, one may define a unitary operator

$$W_\pi \in \mathrm{U}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n), \tag{7.2}$$

for each permutation $\pi \in S_n$, by the action

$$W_\pi(x_1 \otimes \cdots \otimes x_n) = x_{\pi^{-1}(1)} \otimes \cdots \otimes x_{\pi^{-1}(n)} \tag{7.3}$$

for every choice of vectors $x_1, \ldots, x_n \in \mathbb{C}^\Sigma$. The action of the operator W_π, when considered as a channel acting on a state ρ as

$$\rho \mapsto W_\pi \rho W_\pi^*, \tag{7.4}$$

corresponds to the shuffling of the contents of the registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$ that the permutation π describes. Figure 7.1 depicts an example of this action.

One may observe that

$$W_\pi W_\sigma = W_{\pi\sigma} \quad \text{and} \quad W_\pi^{-1} = W_\pi^* = W_{\pi^{-1}} \tag{7.5}$$

for all permutations $\pi, \sigma \in S_n$. Each operator W_π is a permutation operator, in the sense that it is a unitary operator with entries drawn from the set $\{0, 1\}$, and therefore one has

$$\overline{W_\pi} = W_\pi \quad \text{and} \quad W_\pi^{\mathsf{T}} = W_\pi^* \tag{7.6}$$

for every $\pi \in S_n$.

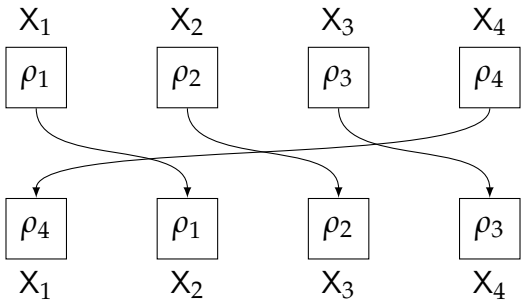


Figure 7.1: The action of the operator W_π on a register $(\mathsf{X}_1, \mathsf{X}_2, \mathsf{X}_3, \mathsf{X}_4)$ when $\pi = (1\ 2\ 3\ 4)$. If the register $(\mathsf{X}_1, \mathsf{X}_2, \mathsf{X}_3, \mathsf{X}_4)$ was initially in the product state $\rho = \rho_1 \otimes \rho_2 \otimes \rho_3 \otimes \rho_4$, and the contents of these registers were permuted according to π as illustrated, the resulting state would then be given by $W_\pi \rho W_\pi^* = \rho_4 \otimes \rho_1 \otimes \rho_2 \otimes \rho_3$. For non-product states, the action of W_π is determined by linearity.

The symmetric subspace

As suggested above, some vectors in $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ are invariant under the action of W_π for every choice of $\pi \in S_n$, and it holds that the set of all such vectors forms a subspace known as the *symmetric subspace*. This subspace will be denoted $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$, which is defined in more precise terms as

$$\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n = \big\{ x \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \,:\, x = W_\pi x \text{ for every } \pi \in S_n \big\}. \tag{7.7}$$

The following proposition states that the projection onto the symmetric subspace is given by a uniform linear combination of the operators W_π, ranging over all $\pi \in S_n$. This serves as a convenient starting point from which other facts regarding the symmetric subspace may be derived.

Proposition 7.1. *Let Σ be an alphabet and let $\mathcal{X}_k = \mathbb{C}^\Sigma$ for $k = 1, \ldots, n$. The projection onto the symmetric subspace $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$ is given by*

$$\Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} = \frac{1}{n!} \sum_{\pi \in S_n} W_\pi. \tag{7.8}$$

Proof. Using the equations (7.5), one may verify directly that the operator

$$\Pi = \frac{1}{n!} \sum_{\pi \in S_n} W_\pi \tag{7.9}$$

is Hermitian and squares to itself, implying that it is a projection operator. It holds that $W_\pi \Pi = \Pi$ for every $\pi \in S_n$, implying that

$$\operatorname{im}(\Pi) \subseteq \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n. \tag{7.10}$$

On the other hand, for every $x \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, it is evident that $\Pi x = x$, implying

$$\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \subseteq \operatorname{im}(\Pi). \tag{7.11}$$

As Π is a projection operator with $\operatorname{im}(\Pi) = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, the proposition is proved. □

Next, an orthonormal basis for the symmetric subspace $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ will be identified, and in the process the dimension of this space will be determined. It is helpful to make use of basic combinatorial concepts for this purpose.

First, for every alphabet Σ and every positive integer n, one defines the set $\operatorname{Bag}(n, \Sigma)$ to be the collection of all functions of the form $\phi : \Sigma \to \mathbb{N}$ (where $\mathbb{N} = \{0, 1, 2, \ldots\}$) possessing the property

$$\sum_{a \in \Sigma} \phi(a) = n. \tag{7.12}$$

Each function $\phi \in \operatorname{Bag}(n, \Sigma)$ may be viewed as describing a *bag* containing a total of n objects, each labeled by a symbol from the alphabet Σ. For each $a \in \Sigma$, the value $\phi(a)$ specifies the number of objects in the bag that are labeled by a. The objects are not considered to be ordered within the bag—it is only the number of objects having each possible label that is indicated by the function ϕ. Equivalently, a function $\phi \in \operatorname{Bag}(n, \Sigma)$ may be interpreted as a description of a multiset of size exactly n with elements drawn from Σ.

An n-tuple $(a_1, \ldots, a_n) \in \Sigma^n$ is *consistent* with a function $\phi \in \operatorname{Bag}(n, \Sigma)$ if and only if

$$\phi(a) = \big|\{k \in \{1, \ldots, n\} \,:\, a = a_k\}\big| \tag{7.13}$$

for every $a \in \Sigma$. In words, $(a_1, \ldots, a_n)$ is consistent with ϕ if and only if $(a_1, \ldots, a_n)$ represents one possible ordering of the elements in the multiset specified by ϕ. For each $\phi \in \operatorname{Bag}(n, \Sigma)$, the set Σ^n_ϕ is defined as the subset of Σ^n containing those elements $(a_1, \ldots, a_n) \in \Sigma^n$ that are consistent with ϕ. This yields a partition of Σ^n, as each n-tuple $(a_1, \ldots, a_n) \in \Sigma^n$ is consistent with precisely one function $\phi \in \operatorname{Bag}(n, \Sigma)$. For any two n-tuples

$$(a_1, \ldots, a_n),\ (b_1, \ldots, b_n) \in \Sigma^n_\phi \tag{7.14}$$

that are consistent with the same function $\phi \in \mathrm{Bag}(n, \Sigma)$, there must exist at least one permutation $\pi \in S_n$ for which

$$(a_1, \ldots, a_n) = \big(b_{\pi(1)}, \ldots, b_{\pi(n)}\big). \tag{7.15}$$

The number of distinct functions $\phi \in \mathrm{Bag}(n, \Sigma)$ is given by the formula

$$|\mathrm{Bag}(n, \Sigma)| = \binom{|\Sigma| + n - 1}{|\Sigma| - 1} \tag{7.16}$$

while the number of distinct n-tuples within the subset Σ_ϕ^n is given by

$$\big|\Sigma_\phi^n\big| = \frac{n!}{\prod_{a \in \Sigma} \big(\phi(a)!\big)} \tag{7.17}$$

for each $\phi \in \mathrm{Bag}(n, \Sigma)$.

An orthonormal basis for the symmetric subspace $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$ may now be obtained through the notions just introduced. Specifically, for each $\phi \in \mathrm{Bag}(n, \Sigma)$, define a unit vector

$$u_\phi = \big|\Sigma_\phi^n\big|^{-\frac{1}{2}} \sum_{(a_1, \ldots, a_n) \in \Sigma_\phi^n} e_{a_1} \otimes \cdots \otimes e_{a_n}. \tag{7.18}$$

As the following proposition establishes, the collection of all such vectors is an orthonormal basis for $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$.

Proposition 7.2. *Let Σ be an alphabet, let n be a positive integer, let $\mathcal{X}_k = \mathbb{C}^\Sigma$ for $k = 1, \ldots, n$, and define $u_\phi \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ for each $\phi \in \mathrm{Bag}(n, \Sigma)$ as in (7.18). It holds that the collection*

$$\big\{u_\phi \,:\, \phi \in \mathrm{Bag}(n, \Sigma)\big\} \tag{7.19}$$

is an orthonormal basis for $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$.

Proof. It is evident that each vector u_ϕ is a unit vector. Moreover, for each choice of $\phi, \psi \in \mathrm{Bag}(n, \Sigma)$ with $\phi \neq \psi$, it holds that

$$\Sigma_\phi^n \cap \Sigma_\psi^n = \varnothing, \tag{7.20}$$

and therefore $\langle u_\phi, u_\psi \rangle = 0$, as each element $(a_1, \ldots, a_n) \in \Sigma^n$ is consistent with precisely one element of $\mathrm{Bag}(n, \Sigma)$. It therefore holds that (7.19) is an

orthonormal set. As each vector u_ϕ is invariant under the action of W_π for every $\pi \in S_n$, it holds that

$$u_\phi \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \tag{7.21}$$

for every $\phi \in \mathrm{Bag}(n, \Sigma)$.

To complete the proof, it remains to prove that the set

$$\{u_\phi \,:\, \phi \in \mathrm{Bag}(n, \Sigma)\} \tag{7.22}$$

spans all of $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$. This fact follows from the observation that, for every n-tuple $(a_1, \ldots, a_n) \in \Sigma^n$, it holds that

$$\begin{aligned} &\Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}(e_{a_1} \otimes \cdots \otimes e_{a_n}) \\ &\quad = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} W_\pi\big(e_{a_1} \otimes \cdots \otimes e_{a_n}\big) = \big|\Sigma_\phi^n\big|^{-\frac{1}{2}} u_\phi, \end{aligned} \tag{7.23}$$

for the unique element $\phi \in \mathrm{Bag}(n, \Sigma)$ with which the n-tuple $(a_1, \ldots, a_n)$ is consistent. □

Corollary 7.3. *Let Σ be an alphabet, let n be a positive integer, and let $\mathcal{X}_k = \mathbb{C}^\Sigma$ for $k = 1, \ldots, n$. It holds that*

$$\dim\big(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n\big) = \binom{|\Sigma| + n - 1}{|\Sigma| - 1}. \tag{7.24}$$

Example 7.4. Suppose $n = 3$ and $\Sigma = \{0, 1\}$. The following four vectors form an orthonormal basis of $\mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \mathcal{X}_3$:

$$\begin{aligned} u_0 &= e_0 \otimes e_0 \otimes e_0 \\ u_1 &= \frac{1}{\sqrt{3}}(e_0 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_0) \\ u_2 &= \frac{1}{\sqrt{3}}(e_0 \otimes e_1 \otimes e_1 + e_1 \otimes e_0 \otimes e_1 + e_1 \otimes e_1 \otimes e_0) \\ u_3 &= e_1 \otimes e_1 \otimes e_1. \end{aligned} \tag{7.25}$$

Tensor power spanning sets for the symmetric subspace

It is evident that the inclusion

$$v^{\otimes n} \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \tag{7.26}$$

holds for every vector $v \in \mathbb{C}^{\Sigma}$. The following theorem demonstrates that the symmetric subspace $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$ is, in fact, spanned by the set of all vectors having this form. This fact remains true when the entries of v are restricted to finite subsets of $\mathbb{C}$, provided that those sets are sufficiently large.

Theorem 7.5. *Let n be a positive integer, let Σ be an alphabet, and let $\mathcal{X}_k = \mathbb{C}^{\Sigma}$ for $k = 1, \ldots, n$. For any set $\mathcal{A} \subseteq \mathbb{C}$ satisfying $|\mathcal{A}| \geq n + 1$ it holds that*

$$\operatorname{span}\Big\{ v^{\otimes n} \,:\, v \in \mathcal{A}^{\Sigma} \Big\} = \mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n. \tag{7.27}$$

Theorem 7.5 can be proved in multiple ways. One proof makes use of the following elementary fact concerning multivariate polynomials.

Lemma 7.6 (Schwartz–Zippel)**.** *Let P be a multivariate complex polynomial in the variables $Z_1, \ldots, Z_m$ that is not identically zero and has total degree at most n, and let $\mathcal{A} \subset \mathbb{C}$ be a nonempty, finite set of complex numbers. It holds that*

$$\big|\big\{(\alpha_1, \ldots, \alpha_m) \in \mathcal{A}^m \,:\, P(\alpha_1, \ldots, \alpha_m) = 0\big\}\big| \leq n|\mathcal{A}|^{m-1}. \tag{7.28}$$

Proof. The lemma is trivial in the case that $|\mathcal{A}| \leq n$, so it will be assumed that $|\mathcal{A}| \geq n + 1$ for the remainder of the proof, which is by induction on m. When $m = 1$, the lemma follows from the fact that a nonzero, univariate polynomial with degree at most n can have at most n roots.

Under the assumption that $m \geq 2$, one may write

$$P(Z_1, \ldots, Z_m) = \sum_{k=0}^{n} Q_k(Z_1, \ldots, Z_{m-1}) Z_m^k, \tag{7.29}$$

for $Q_0, \ldots, Q_n$ being complex polynomials in variables $Z_1, \ldots, Z_{m-1}$, and with the total degree of Q_k being at most $n - k$ for each $k \in \{0, \ldots, n\}$. Fix k to be the largest value in the set $\{0, \ldots, n\}$ for which Q_k is nonzero. Given that P is nonzero, there must exist such a choice of k.

As Q_k has total degree at most $n - k$, it follows from the hypothesis of induction that

$$\begin{aligned} \big|\big\{(\alpha_1, \ldots, \alpha_{m-1}) \in \mathcal{A}^{m-1} \,:\, Q_k(\alpha_1, \ldots, \alpha_{m-1}) \neq 0\big\}\big| \\ \geq |\mathcal{A}|^{m-1} - (n-k)|\mathcal{A}|^{m-2}. \end{aligned} \tag{7.30}$$

For each choice of $(\alpha_1, \ldots, \alpha_{m-1}) \in \mathcal{A}^{m-1}$ for which $Q_k(\alpha_1, \ldots, \alpha_{m-1}) \neq 0$, it holds that

$$P(\alpha_1, \ldots, \alpha_{m-1}, Z_m) = \sum_{j=0}^{n} Q_j(\alpha_1, \ldots, \alpha_{m-1}) Z_m^j \tag{7.31}$$

is a univariate polynomial of degree k in the variable Z_m, implying that there must exist at least $|\mathcal{A}| - k$ choices of $\alpha_m \in \mathcal{A}$ for which

$$P(\alpha_1, \ldots, \alpha_m) \neq 0. \tag{7.32}$$

It follows that there are at least

$$\big(|\mathcal{A}|^{m-1} - (n-k)|\mathcal{A}|^{m-2}\big)\big(|\mathcal{A}| - k\big) \geq |\mathcal{A}|^m - n|\mathcal{A}|^{m-1} \tag{7.33}$$

distinct m-tuples $(\alpha_1, \ldots, \alpha_m) \in \mathcal{A}^m$ for which $P(\alpha_1, \ldots, \alpha_m) \neq 0$, which completes the proof of the lemma. □

Remark 7.7. Although it is irrelevant to its use in proving Theorem 7.5, one may observe that Lemma 7.6 holds for P being a multivariate polynomial over any field, not just the field of complex numbers. This fact is established by the proof above, which has not used properties of the complex numbers that do not hold for arbitrary fields.

Proof of Theorem 7.5. For every choice of a permutation $\pi \in S_n$ and a vector $v \in \mathbb{C}^\Sigma$, it holds that

$$W_\pi\big(v^{\otimes n}\big) = v^{\otimes n}. \tag{7.34}$$

It follows that $v^{\otimes n} \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, and therefore

$$\operatorname{span}\Big\{v^{\otimes n} \,:\, v \in \mathcal{A}^\Sigma\Big\} \subseteq \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n. \tag{7.35}$$

To prove the reverse inclusion, let $w \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ be any nonzero vector, and write

$$w = \sum_{\phi \in \operatorname{Bag}(n,\Sigma)} \alpha_\phi u_\phi, \tag{7.36}$$

for some collection of coefficients

$$\{\alpha_\phi \,:\, \phi \in \operatorname{Bag}(n,\Sigma)\} \subset \mathbb{C}, \tag{7.37}$$

with each vector u_ϕ being defined as in (7.18). It will be proved that

$$\langle w, v^{\otimes n}\rangle \neq 0 \tag{7.38}$$

for at least one choice of a vector $v \in \mathcal{A}^\Sigma$. The required inclusion follows from this fact, for if the containment (7.35) were proper, it would be possible to choose $w \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ that is orthogonal to $v^{\otimes n}$ for every $v \in \mathcal{A}^\Sigma$.

For the remainder of the proof it will be assumed that $\mathcal{A}$ is a finite set, which causes no loss of generality—for if $\mathcal{A}$ were infinite, one could restrict their attention to an arbitrary finite subset of $\mathcal{A}$ having size at least $n+1$, yielding the desired inclusion.

Define a multivariate polynomial

$$Q = \sum_{\phi \in \mathrm{Bag}(n,\Sigma)} \overline{\alpha_\phi} \sqrt{|\Sigma_\phi^n|} \prod_{a \in \Sigma} Z_a^{\phi(a)} \tag{7.39}$$

in a collection of variables $\{Z_a \,:\, a \in \Sigma\}$. As the monomials

$$\prod_{a \in \Sigma} Z_a^{\phi(a)} \tag{7.40}$$

are distinct as ϕ ranges over the elements of $\mathrm{Bag}(n,\Sigma)$, with each monomial having total degree n, it follows that Q is a nonzero polynomial with total degree n. A calculation reveals that

$$Q(v) = \langle w, v^{\otimes n} \rangle \tag{7.41}$$

for every vector $v \in \mathbb{C}^\Sigma$, where $Q(v)$ refers to the complex number obtained by the substitution of the value $v(a)$ for the variable Z_a in Q for each $a \in \Sigma$. As Q is a nonzero multivariate polynomial with total degree n, it follows from Lemma 7.6 that $Q(v) = 0$ for at most

$$n|\mathcal{A}|^{|\Sigma|-1} < |\mathcal{A}|^{|\Sigma|} \tag{7.42}$$

choices of vectors $v \in \mathcal{A}^\Sigma$, implying that there exists at least one vector $v \in \mathcal{A}^\Sigma$ for which $\langle w, v^{\otimes n} \rangle \neq 0$, completing the proof. □

The anti-symmetric subspace

Along similar lines to the symmetric subspace $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$ of the tensor product space $\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, one may define the *anti-symmetric subspace* of the same tensor product space as

$$\begin{aligned} &\mathcal{X}_1 \circledast \cdots \circledast \mathcal{X}_n \\ &\quad = \bigl\{ x \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \,:\, W_\pi x = \mathrm{sign}(\pi) x \text{ for every } \pi \in S_n \bigr\}. \end{aligned} \tag{7.43}$$

The short discussion on the anti-symmetric subspace that follows may, for the most part, be considered as an aside; with the exception of the case in

which $n = 2$, the anti-symmetric subspace does not play a significant role elsewhere in this book. It is, nevertheless, natural to consider this subspace along side of the symmetric subspace. The following propositions establish a few basic facts about the anti-symmetric subspace.

Proposition 7.8. *Let Σ be an alphabet and let $\mathcal{X}_k = \mathbb{C}^\Sigma$ for $k = 1, \ldots, n$. The projection onto the anti-symmetric subspace $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$ is given by*

$$\Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} = \frac{1}{n!} \sum_{\pi \in S_n} \operatorname{sign}(\pi) W_\pi. \tag{7.44}$$

Proof. The proof is similar to the proof of Proposition 7.1. Using (7.5), along with the fact that $\operatorname{sign}(\pi)\operatorname{sign}(\sigma) = \operatorname{sign}(\pi\sigma)$ for every choice of $\pi, \sigma \in S_n$, it may be verified that the operator

$$\Pi = \frac{1}{n!} \sum_{\pi \in S_n} \operatorname{sign}(\pi) W_\pi \tag{7.45}$$

is Hermitian and squares to itself, implying that it is a projection operator. For every $\pi \in S_n$ it holds that

$$W_\pi \Pi = \operatorname{sign}(\pi) \Pi, \tag{7.46}$$

from which it follows that

$$\operatorname{im}(\Pi) \subseteq \mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n. \tag{7.47}$$

For every vector $x \in \mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$, it holds that $\Pi x = x$, implying that

$$\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n \subseteq \operatorname{im}(\Pi). \tag{7.48}$$

As Π is a projection operator with $\operatorname{im}(\Pi) = \mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$, the proposition is proved. □

When constructing an orthonormal basis of the anti-symmetric subspace $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$, it is convenient to assume that a total ordering of Σ has been fixed. For every n-tuple $(a_1, \ldots, a_n) \in \Sigma^n$ for which $a_1 < \cdots < a_n$, define a vector

$$u_{a_1, \ldots, a_n} = \frac{1}{\sqrt{n!}} \sum_{\pi \in S_n} \operatorname{sign}(\pi) W_\pi (e_{a_1} \otimes \cdots \otimes e_{a_n}). \tag{7.49}$$

Proposition 7.9. *Let n be a positive integer with $n \geq 2$, let Σ be an alphabet, let $\mathcal{X}_k = \mathbb{C}^{\Sigma}$ for $k = 1, \ldots, n$, and define $u_{a_1,\ldots,a_n} \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ for each n-tuple $(a_1, \ldots, a_n) \in \Sigma^n$ satisfying $a_1 < \cdots < a_n$ as in (7.49). The collection*

$$\big\{ u_{a_1,\ldots,a_n} \,:\, (a_1, \ldots, a_n) \in \Sigma^n,\ a_1 < \cdots < a_n \big\} \tag{7.50}$$

is an orthonormal basis for $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$.

Proof. Each vector $u_{a_1,\ldots,a_n}$ is evidently a unit vector, and is contained in the space $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$. For distinct n-tuples $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ with $a_1 < \cdots < a_n$ and $b_1 < \cdots < b_n$ it holds that

$$\langle u_{a_1,\ldots,a_n}, u_{b_1,\ldots,b_n} \rangle = 0, \tag{7.51}$$

as these vectors are linear combinations of disjoint sets of standard basis vectors. It therefore remains to prove that the collection (7.50) spans all of $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$.

For any choice of distinct indices $j, k \in \{1, \ldots, n\}$, and for $(j\,k) \in S_n$ being the permutation that swaps j and k, leaving all other elements of $\{1, \ldots, n\}$ fixed, one has

$$W_{(j\,k)} \Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} = -\Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} = \Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} W_{(j\,k)}. \tag{7.52}$$

Consequently, for any choice of an n-tuple $(a_1, \ldots, a_n) \in \Sigma^n$ for which there exist distinct indices $j, k \in \{1, \ldots, n\}$ for which $a_j = a_k$, it holds that

$$\begin{aligned} \Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} (e_{a_1} \otimes \cdots \otimes e_{a_n}) &= \Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} W_{(j\,k)} (e_{a_1} \otimes \cdots \otimes e_{a_n}) \\ &= -\Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} (e_{a_1} \otimes \cdots \otimes e_{a_n}), \end{aligned} \tag{7.53}$$

and therefore

$$\Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} (e_{a_1} \otimes \cdots \otimes e_{a_n}) = 0. \tag{7.54}$$

On the other hand, for $(a_1, \ldots, a_n) \in \Sigma^n$ being an n-tuple for which $a_1, \ldots, a_n$ are distinct elements of Σ, it must hold that

$$(a_1, \ldots, a_n) = \pi(b_1, \ldots, b_n) \tag{7.55}$$

for some choice of a permutation $\pi \in S_n$ and an n-tuple $(b_1, \ldots, b_n) \in \Sigma^n$ satisfying $b_1 < \cdots < b_n$. One therefore has

$$\begin{aligned} \Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} (e_{a_1} \otimes \cdots \otimes e_{a_n}) &= \Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} W_{\pi} (e_{b_1} \otimes \cdots \otimes e_{b_n}) \\ = \operatorname{sign}(\pi) \Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n} (e_{b_1} \otimes \cdots \otimes e_{b_n}) &= \frac{\operatorname{sign}(\pi)}{\sqrt{n!}} u_{b_1,\ldots,b_n}. \end{aligned} \tag{7.56}$$

It therefore holds that

$$\begin{aligned}&\operatorname{im}\bigl(\Pi_{\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n}\bigr)\\ &\quad \subseteq \operatorname{span}\bigl\{u_{a_1,\ldots,a_n} \,:\, (a_1,\ldots,a_n) \in \Sigma^n,\ a_1 < \cdots < a_n\bigr\},\end{aligned} \tag{7.57}$$

which completes the proof. □

By the previous proposition, one has that the dimension of the anti-symmetric subspace is equal to the number of n-tuples $(a_1,\ldots,a_n) \in \Sigma^n$ satisfying $a_1 < \cdots < a_n$. This number is equal to the number of subsets of Σ having n elements.

Corollary 7.10. *Let n be a positive integer, let Σ be an alphabet, and let $\mathcal{X}_k = \mathbb{C}^\Sigma$ for $k = 1,\ldots,n$. It holds that*

$$\dim\bigl(\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n\bigr) = \binom{|\Sigma|}{n}. \tag{7.58}$$

7.1.2 The algebra of permutation-invariant operators

As it was defined above, the symmetric subspace $\mathcal{X}_1 \circledcirc \cdots \circledcirc \mathcal{X}_n$ includes those vectors $x \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ that are invariant under the action of W_π for each $\pi \in S_n$. One may consider a similar notion for operators, with the action $x \mapsto W_\pi x$ being replaced by the action

$$X \mapsto W_\pi X W_\pi^* \tag{7.59}$$

for each $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$. The notation $\mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n)$ will be used to denote the set of operators X that are invariant under this action:

$$\begin{aligned}&\mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n)\\ &\quad = \bigl\{X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n) \,:\, X = W_\pi X W_\pi^* \text{ for all } \pi \in S_n\bigr\}.\end{aligned} \tag{7.60}$$

Density operator elements of the set $\mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n)$ represent states of the compound register $(\mathsf{X}_1,\ldots,\mathsf{X}_n)$ that are invariant under permutations of the registers $\mathsf{X}_1,\ldots,\mathsf{X}_n$. Such states are said to be *exchangeable*. Algebraic properties of the set $\mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n)$, along with a relationship between exchangeable states and permutation-invariant vectors, are described in the subsections that follow.

Vector space structure of the permutation-invariant operators

The notation $\mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n)$ is a natural choice for the space of all permutation-invariant operators—if one regards $\mathrm{L}(\mathcal{X}_1), \ldots, \mathrm{L}(\mathcal{X}_n)$ as vector spaces, ignoring the operator structure that is defined on these sets, then $\mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n)$ indeed coincides with the symmetric subspace of the tensor product space $\mathrm{L}(\mathcal{X}_1) \otimes \cdots \otimes \mathrm{L}(\mathcal{X}_n)$. The next proposition formalizes this connection and states some immediate consequences from the results of the previous section.

Proposition 7.11. *Let n be a positive integer, let Σ be an alphabet, let $\mathcal{X}_k = \mathbb{C}^\Sigma$ for $k = 1, \ldots, n$, and let $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$. The following statements are equivalent:*

1. $X \in \mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n)$.
2. *For $V \in \mathrm{U}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \otimes \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, (\mathcal{X}_1 \otimes \mathcal{X}_1) \otimes \cdots \otimes (\mathcal{X}_n \otimes \mathcal{X}_n))$ being the isometry defined by the equation*

$$V \operatorname{vec}(Y_1 \otimes \cdots \otimes Y_n) = \operatorname{vec}(Y_1) \otimes \cdots \otimes \operatorname{vec}(Y_n) \tag{7.61}$$

   holding for all $Y_1 \in \mathrm{L}(\mathcal{X}_1), \ldots, Y_n \in \mathrm{L}(\mathcal{X}_n)$, one has that

$$V \operatorname{vec}(X) \in (\mathcal{X}_1 \otimes \mathcal{X}_1) \circledcirc \cdots \circledcirc (\mathcal{X}_n \otimes \mathcal{X}_n). \tag{7.62}$$

3. $X \in \operatorname{span}\{Y^{\otimes n} : Y \in \mathrm{L}(\mathbb{C}^\Sigma)\}$.

Proof. For each permutation $\pi \in S_n$, let

$$U_\pi \in \mathrm{U}\big((\mathcal{X}_1 \otimes \mathcal{X}_1) \otimes \cdots \otimes (\mathcal{X}_n \otimes \mathcal{X}_n)\big) \tag{7.63}$$

be the unitary operator defined by the equation

$$U_\pi(w_1 \otimes \cdots \otimes w_n) = w_{\pi^{-1}(1)} \otimes \cdots \otimes w_{\pi^{-1}(n)} \tag{7.64}$$

holding for all vectors $w_1 \in \mathcal{X}_1 \otimes \mathcal{X}_1$, …, $w_n \in \mathcal{X}_n \otimes \mathcal{X}_n$. Each operator U_π is analogous to W_π, as defined in (7.3), but with the space $\mathcal{X}_k$ replaced by $\mathcal{X}_k \otimes \mathcal{X}_k$ for each $k \in \{1, \ldots, n\}$. It holds that

$$U_\pi = V(W_\pi \otimes W_\pi)V^* \tag{7.65}$$

for every $\pi \in S_n$, from which one may conclude that the first and second statements are equivalent.

Theorem 7.5 implies that

$$V\operatorname{vec}(X) \in (\mathcal{X}_1 \otimes \mathcal{X}_1) \circledcirc \cdots \circledcirc (\mathcal{X}_n \otimes \mathcal{X}_n) \tag{7.66}$$

if and only if

$$V\operatorname{vec}(X) \in \operatorname{span}\bigl\{\operatorname{vec}(Y)^{\otimes n} \,:\, Y \in \mathrm{L}\bigl(\mathbb{C}^{\Sigma}\bigr)\bigr\}. \tag{7.67}$$

The containment (7.67) is equivalent to

$$\operatorname{vec}(X) \in \operatorname{span}\bigl\{\operatorname{vec}\bigl(Y^{\otimes n}\bigr) \,:\, Y \in \mathrm{L}\bigl(\mathbb{C}^{\Sigma}\bigr)\bigr\}, \tag{7.68}$$

which in turn is equivalent to

$$X \in \operatorname{span}\bigl\{Y^{\otimes n} \,:\, Y \in \mathrm{L}\bigl(\mathbb{C}^{\Sigma}\bigr)\bigr\}. \tag{7.69}$$

The second and third statements are therefore equivalent. □

Theorem 7.12. *Let n be a positive integer, let Σ be an alphabet, and let $\mathcal{X}_k = \mathbb{C}^{\Sigma}$ for $k = 1, \ldots, n$. It holds that*

$$\mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n) = \operatorname{span}\bigl\{U^{\otimes n} \,:\, U \in \mathrm{U}\bigl(\mathbb{C}^{\Sigma}\bigr)\bigr\}. \tag{7.70}$$

Proof. Let

$$D = \sum_{a \in \Sigma} u(a) E_{a,a} \tag{7.71}$$

be a diagonal operator, for an arbitrary choice of $u \in \mathbb{C}^{\Sigma}$. By Theorem 7.5 it follows that

$$u^{\otimes n} \in \operatorname{span}\bigl\{v^{\otimes n} \,:\, v \in \mathbb{T}^{\Sigma}\bigr\}, \tag{7.72}$$

for $\mathbb{T} = \bigl\{\alpha \in \mathbb{C} \,:\, |\alpha| = 1\bigr\}$ denoting the group of complex units, so that

$$u^{\otimes n} = \sum_{b \in \Gamma} \beta_b v_b^{\otimes n} \tag{7.73}$$

for some choice of an alphabet Γ, vectors $\{v_b \,:\, b \in \Gamma\} \subset \mathbb{T}^{\Sigma}$, and complex numbers $\{\beta_b \,:\, b \in \Gamma\} \subset \mathbb{C}$. It follows that

$$D^{\otimes n} = \sum_{b \in \Gamma} \beta_b U_b^{\otimes n} \tag{7.74}$$

for $U_b \in \mathrm{U}(\mathbb{C}^{\Sigma})$ being the unitary operator defined as

$$U_b = \sum_{a \in \Sigma} v_b(a) E_{a,a} \tag{7.75}$$

for each $b \in \Gamma$.

Now, for an arbitrary operator $A \in \mathrm{L}\big(\mathbb{C}^{\Sigma}\big)$, one may write $A = VDW$ for $V, W \in \mathrm{U}\big(\mathbb{C}^{\Sigma}\big)$ being unitary operators and $D \in \mathrm{L}\big(\mathbb{C}^{\Sigma}\big)$ being a diagonal operator, by means of the singular value theorem (Theorem 1.6). Invoking the argument above, one may assume that (7.74) holds, and therefore

$$A^{\otimes n} = \sum_{b \in \Gamma} \beta_b (V U_b W)^{\otimes n}, \tag{7.76}$$

for some choice of an alphabet Γ, complex numbers $\{\beta_b \,:\, b \in \Gamma\} \subset \mathbb{C}$, and diagonal unitary operators $\{U_b \,:\, b \in \Gamma\}$. As VU_bW is unitary for each $b \in \Gamma$, it holds that

$$A^{\otimes n} \in \operatorname{span}\big\{U^{\otimes n} \,:\, U \in \mathrm{U}\big(\mathbb{C}^{\Sigma}\big)\big\}, \tag{7.77}$$

and therefore

$$\mathrm{L}(\mathcal{X}_1) \otimes\!\!\!\!\bigcirc \cdots \otimes\!\!\!\!\bigcirc \mathrm{L}(\mathcal{X}_n) \subseteq \operatorname{span}\big\{U^{\otimes n} \,:\, U \in \mathrm{U}\big(\mathbb{C}^{\Sigma}\big)\big\}. \tag{7.78}$$

The reverse containment is immediate, so the theorem is proved. □

Symmetric purifications of exchangeable density operators

A density operator $\rho \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ is exchangeable if and only if $\rho \in \mathrm{L}(\mathcal{X}_1) \otimes\!\!\!\!\bigcirc \cdots \otimes\!\!\!\!\bigcirc \mathrm{L}(\mathcal{X}_n)$, which is equivalent to

$$\rho = W_\pi \rho W_\pi^* \tag{7.79}$$

for every permutation $\pi \in S_n$. In operational terms, an exchangeable state ρ of a compound register $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$, for n identical registers $\mathsf{X}_1, \ldots, \mathsf{X}_n$, is one that does not change if the contents of these n registers are permuted in an arbitrary way.

For every symmetric unit vector $u \in \mathcal{X}_1 \otimes\!\!\!\!\bigcirc \cdots \otimes\!\!\!\!\bigcirc \mathcal{X}_n$, one has that the pure state uu^* is exchangeable, and naturally any convex combination of such states must be exchangeable as well. In general, this does not exhaust all possible exchangeable states. For instance, the completely mixed state in $\mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ is exchangeable, but the image of the density operator corresponding to this state is generally not contained within the symmetric subspace.

There is, nevertheless, an interesting relationship between exchangeable states and symmetric pure states, which is that every exchangeable state can be purified in such a way that its purification lies within a larger symmetric subspace, in the sense described by the following theorem.

Theorem 7.13. *Let n be a positive integer, let $\mathsf{X}_1, \ldots, \mathsf{X}_n$ be registers having the same classical state set Σ, let $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ be registers having the same classical state set Γ, and assume $|\Gamma| \geq |\Sigma|$. Suppose further that $\rho \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ is an exchangeable density operator. There exists a unit vector*

$$u \in (\mathcal{X}_1 \otimes \mathcal{Y}_1) \circledcirc \cdots \circledcirc (\mathcal{X}_n \otimes \mathcal{Y}_n) \tag{7.80}$$

such that

$$(uu^*)[\mathsf{X}_1, \ldots, \mathsf{X}_n] = \rho. \tag{7.81}$$

Proof. Let $A \in \mathrm{U}(\mathbb{C}^\Sigma, \mathbb{C}^\Gamma)$ be an arbitrarily chosen isometry, which one may regard as an element of $\mathrm{U}(\mathcal{X}_k, \mathcal{Y}_k)$ for any choice of $k \in \{1, \ldots, n\}$. Also let

$$\begin{aligned} V \in \mathrm{U}\big((\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n) \otimes (\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_n), \\ (\mathcal{X}_1 \otimes \mathcal{Y}_1) \otimes \cdots \otimes (\mathcal{X}_n \otimes \mathcal{Y}_n)\big) \end{aligned} \tag{7.82}$$

be the isometry defined by the equation

$$V \operatorname{vec}(B_1 \otimes \cdots \otimes B_n) = \operatorname{vec}(B_1) \otimes \cdots \otimes \operatorname{vec}(B_n), \tag{7.83}$$

holding for every choice of $B_1 \in \mathrm{L}(\mathcal{Y}_1, \mathcal{X}_1), \ldots, B_n \in \mathrm{L}(\mathcal{Y}_n, \mathcal{X}_n)$. Equivalently, this isometry is defined by the equation

$$\begin{aligned} V((x_1 \otimes \cdots \otimes x_n) \otimes (y_1 \otimes \cdots \otimes y_n)) \\ = (x_1 \otimes y_1) \otimes \cdots \otimes (x_n \otimes y_n), \end{aligned} \tag{7.84}$$

holding for all vectors $x_1 \in \mathcal{X}_1, \ldots, x_n \in \mathcal{X}_n$ and $y_1 \in \mathcal{Y}_1, \ldots, y_n \in \mathcal{Y}_n$.

Consider the vector

$$u = V \operatorname{vec}\big(\sqrt{\rho}(A^* \otimes \cdots \otimes A^*)\big) \in (\mathcal{X}_1 \otimes \mathcal{Y}_1) \otimes \cdots \otimes (\mathcal{X}_n \otimes \mathcal{Y}_n). \tag{7.85}$$

A calculation reveals that

$$(uu^*)[\mathsf{X}_1, \ldots, \mathsf{X}_n] = \rho, \tag{7.86}$$

and so it remains to prove that u is symmetric. Because ρ is exchangeable, one has

$$\big(W_\pi \sqrt{\rho} W_\pi^*\big)^2 = W_\pi \rho W_\pi^* = \rho \tag{7.87}$$

for every permutation $\pi \in \mathcal{S}_n$, and therefore

$$W_\pi \sqrt{\rho} W_\pi^* = \sqrt{\rho} \tag{7.88}$$

by the uniqueness of the square root. By Proposition 7.11, it therefore holds that

$$\sqrt{\rho} \in \operatorname{span}\bigl\{Y^{\otimes n} \,:\, Y \in \mathrm{L}(\mathbb{C}^{\Sigma})\bigr\}. \tag{7.89}$$

Consequently, one has

$$u \in \operatorname{span}\Bigl\{V \operatorname{vec}\Bigl(\bigl(YA^*\bigr)^{\otimes n}\Bigr) \,:\, Y \in \mathrm{L}(\mathbb{C}^{\Sigma})\Bigr\}, \tag{7.90}$$

and therefore

$$u \in \operatorname{span}\Bigl\{\operatorname{vec}\bigl(YA^*\bigr)^{\otimes n} \,:\, Y \in \mathrm{L}(\mathbb{C}^{\Sigma})\Bigr\}. \tag{7.91}$$

From this containment it is evident that

$$u \in (\mathcal{X}_1 \otimes \mathcal{Y}_1) \circledcirc \cdots \circledcirc (\mathcal{X}_n \otimes \mathcal{Y}_n), \tag{7.92}$$

which completes the proof. □

Von Neumann's double commutant theorem

To establish further properties of the set $\mathrm{L}(\mathcal{X}_1) \circledcirc \cdots \circledcirc \mathrm{L}(\mathcal{X}_n)$, particularly those relating to the operator structure of its elements, it is convenient to make use of a theorem known as *von Neumann's double commutant theorem*. This theorem is stated below, and its proof will make use of the following lemma.

Lemma 7.14. *Let $\mathcal{X}$ be a complex Euclidean space, let $\mathcal{V} \subseteq \mathcal{X}$ be a subspace of $\mathcal{X}$, and let $A \in \mathrm{L}(\mathcal{X})$ be an operator. The following two statements are equivalent:*

1. *It holds that both $A\mathcal{V} \subseteq \mathcal{V}$ and $A^*\mathcal{V} \subseteq \mathcal{V}$.*
2. *It holds that $[A, \Pi_{\mathcal{V}}] = 0$.*

Proof. Assume first that statement 2 holds. If two operators commute, then their adjoints must also commute, and so one has the following for every vector $v \in \mathcal{V}$:

$$\begin{aligned} Av &= A\Pi_{\mathcal{V}}v = \Pi_{\mathcal{V}}Av \in \mathcal{V}, \\ A^*v &= A^*\Pi_{\mathcal{V}}v = \Pi_{\mathcal{V}}A^*v \in \mathcal{V}. \end{aligned} \tag{7.93}$$

It has been proved that statement 2 implies statement 1.

Now assume statement 1 holds. For every $v \in \mathcal{V}$, one has

$$\Pi_{\mathcal{V}}Av = Av = A\Pi_{\mathcal{V}}v, \tag{7.94}$$

by virtue of the fact that $Av \in \mathcal{V}$. For every $w \in \mathcal{V}^{\perp}$, it must hold that

$$\langle v, Aw \rangle = \langle A^* v, w \rangle = 0 \tag{7.95}$$

for for every $v \in \mathcal{V}$, following from the assumption $A^* v \in \mathcal{V}$, and therefore $Aw \in \mathcal{V}^{\perp}$. Consequently,

$$\Pi_{\mathcal{V}} A w = 0 = A \Pi_{\mathcal{V}} w. \tag{7.96}$$

As every vector $u \in \mathcal{X}$ may be written as $u = v + w$ for some choice of $v \in \mathcal{V}$ and $w \in \mathcal{V}^{\perp}$, equations (7.94) and (7.96) imply

$$\Pi_{\mathcal{V}} A u = A \Pi_{\mathcal{V}} u \tag{7.97}$$

for every vector $u \in \mathcal{X}$, and therefore $\Pi_{\mathcal{V}} A = A \Pi_{\mathcal{V}}$. It has been proved that statement 1 implies statement 2, which completes the proof. □

Theorem 7.15 (Von Neumann's double commutant theorem)**.** *Let $\mathcal{A} \subseteq \mathrm{L}(\mathcal{X})$ be a self-adjoint, unital subalgebra of $\mathrm{L}(\mathcal{X})$, for $\mathcal{X}$ being a complex Euclidean space. It holds that*

$$\mathrm{comm}(\mathrm{comm}(\mathcal{A})) = \mathcal{A}. \tag{7.98}$$

Proof. It is immediate from the definition of the commutant that

$$\mathcal{A} \subseteq \mathrm{comm}(\mathrm{comm}(\mathcal{A})), \tag{7.99}$$

and so it remains to prove the reverse inclusion.

The key idea of the proof will be to consider the algebra $\mathrm{L}(\mathcal{X} \otimes \mathcal{X})$, and to make use of its relationships with $\mathrm{L}(\mathcal{X})$. Define $\mathcal{B} \subseteq \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$ as

$$\mathcal{B} = \{X \otimes \mathbb{1} \,:\, X \in \mathcal{A}\}, \tag{7.100}$$

and let Σ be the alphabet for which $\mathcal{X} = \mathbb{C}^{\Sigma}$. Every operator $Y \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$ may be written as

$$Y = \sum_{a,b \in \Sigma} Y_{a,b} \otimes E_{a,b} \tag{7.101}$$

for a unique choice of operators $\{Y_{a,b} \,:\, a, b \in \Sigma\} \subset \mathrm{L}(\mathcal{X})$. The condition

$$Y(X \otimes \mathbb{1}) = (X \otimes \mathbb{1})Y, \tag{7.102}$$

for any operator $X \in \mathrm{L}(\mathcal{X})$ and any operator Y having the form (7.101), is equivalent to $[Y_{a,b}, X] = 0$ for every choice of $a, b \in \Sigma$, and so it follows that

$$\operatorname{comm}(\mathcal{B}) = \left\{ \sum_{a,b \in \Sigma} Y_{a,b} \otimes E_{a,b} \,:\, \{Y_{a,b} \,:\, a, b \in \Sigma\} \subset \operatorname{comm}(\mathcal{A}) \right\}. \tag{7.103}$$

For a given operator $X \in \operatorname{comm}(\operatorname{comm}(\mathcal{A}))$, it is therefore evident that

$$X \otimes \mathbb{1} \in \operatorname{comm}\big(\operatorname{comm}\big(\mathcal{B}\big)\big). \tag{7.104}$$

Now, define a subspace $\mathcal{V} \subseteq \mathcal{X} \otimes \mathcal{X}$ as

$$\mathcal{V} = \{\operatorname{vec}(X) \,:\, X \in \mathcal{A}\}, \tag{7.105}$$

and let $X \in \mathcal{A}$ be chosen arbitrarily. It holds that

$$(X \otimes \mathbb{1})\mathcal{V} \subseteq \mathcal{V}, \tag{7.106}$$

owing to the fact that $\mathcal{A}$ is an algebra. As $\mathcal{A}$ is self-adjoint, it follows that $X^* \in \mathcal{A}$, and therefore

$$(X^* \otimes \mathbb{1})\mathcal{V} \subseteq \mathcal{V}. \tag{7.107}$$

Lemma 7.14 therefore implies that

$$[X \otimes \mathbb{1}, \Pi_{\mathcal{V}}] = 0. \tag{7.108}$$

As $X \in \mathcal{A}$ was chosen arbitrarily, it follows that $\Pi_{\mathcal{V}} \in \operatorname{comm}(\mathcal{B})$.

Finally, let $X \in \operatorname{comm}(\operatorname{comm}(\mathcal{A}))$ be chosen arbitrarily. As was argued above, the inclusion (7.104) therefore holds, from which the commutation relation (7.108) follows. The reverse implication of Lemma 7.14 implies the containment (7.106). In particular, given that $\mathcal{A}$ is unital, one has $\operatorname{vec}(\mathbb{1}) \in \mathcal{V}$, and therefore

$$\operatorname{vec}(X) = (X \otimes \mathbb{1}) \operatorname{vec}(\mathbb{1}) \in \mathcal{V}, \tag{7.109}$$

which implies $X \in \mathcal{A}$. The containment

$$\operatorname{comm}(\operatorname{comm}(\mathcal{A})) \subseteq \mathcal{A} \tag{7.110}$$

has therefore been proved, which completes the proof. □

Operator structure of the permutation-invariant operators

With von Neumann's double commutant theorem in hand, one is prepared to prove the following fundamental theorem, which concerns the operator structure of the set $\mathrm{L}(\mathcal{X}_1) \otimes \cdots \otimes \mathrm{L}(\mathcal{X}_n)$.

Theorem 7.16. *Let n be a positive integer, let Σ be an alphabet, let $\mathcal{X}_k = \mathbb{C}^{\Sigma}$ for $k = 1, \ldots, n$, and let $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ be an operator. The following statements are equivalent:*

1. *It holds that $[X, Y^{\otimes n}] = 0$ for all choices of $Y \in \mathrm{L}(\mathbb{C}^{\Sigma})$.*
2. *It holds that $[X, U^{\otimes n}] = 0$ for all choices of $U \in \mathrm{U}(\mathbb{C}^{\Sigma})$.*
3. *It holds that*

$$X = \sum_{\pi \in S_n} u(\pi) W_\pi \tag{7.111}$$

for some choice of a vector $u \in \mathbb{C}^{S_n}$.

Proof. By Proposition 7.11 and Theorem 7.12, together with the bilinearity of the Lie bracket, one has that the first and second statements are equivalent to the inclusion

$$X \in \operatorname{comm}\bigl(\mathrm{L}(\mathcal{X}_1) \otimes \cdots \otimes \mathrm{L}(\mathcal{X}_n)\bigr). \tag{7.112}$$

For the set $\mathcal{A} \subseteq \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ defined as

$$\mathcal{A} = \left\{ \sum_{\pi \in S_n} u(\pi) W_\pi \,:\, u \in \mathbb{C}^{S_n} \right\}, \tag{7.113}$$

one has that the third statement is equivalent to the inclusion $X \in \mathcal{A}$. To prove the theorem, it therefore suffices to demonstrate that

$$\mathcal{A} = \operatorname{comm}\bigl(\mathrm{L}(\mathcal{X}_1) \otimes \cdots \otimes \mathrm{L}(\mathcal{X}_n)\bigr). \tag{7.114}$$

For any operator $Z \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$, it is evident from an inspection of (7.60) that $Z \in \mathrm{L}(\mathcal{X}_1) \otimes \cdots \otimes \mathrm{L}(\mathcal{X}_n)$ if and only if $[Z, W_\pi] = 0$ for each $\pi \in S_n$. Again using the bilinearity of the Lie bracket, it follows that

$$\mathrm{L}(\mathcal{X}_1) \otimes \cdots \otimes \mathrm{L}(\mathcal{X}_n) = \operatorname{comm}(\mathcal{A}). \tag{7.115}$$

Finally, one observes that the set $\mathcal{A}$ forms a self-adjoint, unital subalgebra of $\mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$. By Theorem 7.15, one has

$$\operatorname{comm}\bigl(\mathrm{L}(\mathcal{X}_1) \otimes \cdots \otimes \mathrm{L}(\mathcal{X}_n)\bigr) = \operatorname{comm}(\operatorname{comm}(\mathcal{A})) = \mathcal{A}, \tag{7.116}$$

which establishes the relation (7.114), and therefore completes the proof. □

7.2 Unitarily invariant probability measures

Two probability measures having fundamental importance in the theory of quantum information are introduced in the present section: the *uniform spherical measure*, defined on the unit sphere $\mathcal{S}(\mathcal{X})$, and the *Haar measure*, defined on the set of unitary operators $\mathrm{U}(\mathcal{X})$, for every complex Euclidean space $\mathcal{X}$. These measures are closely connected, and may both be defined in simple and concrete terms based on the standard Gaussian measure on the real line (q.v. Section 1.2.1).

7.2.1 Uniform spherical measure and Haar measure basics

Definitions and basic properties of the uniform spherical measure and Haar measure are discussed below, starting with the uniform spherical measure.

Uniform spherical measure

Intuitively speaking, the uniform spherical measure provides a formalism through which one may consider a probability distribution over vectors in a complex Euclidean space that is uniform over the unit sphere. In more precise terms, the uniform spherical measure is a probability measure μ, defined on the Borel subsets of the unit sphere $\mathcal{S}(\mathcal{X})$ of a complex Euclidean space $\mathcal{X}$, that is invariant under the action of every unitary operator:

$$\mu(\mathcal{A}) = \mu(U\mathcal{A}) \tag{7.117}$$

for every $\mathcal{A} \in \mathrm{Borel}(\mathcal{S}(\mathcal{X}))$ and $U \in \mathrm{U}(\mathcal{X})$.[1] One concrete way of defining such a measure is as follows.

Definition 7.17. Let Σ be an alphabet, let $\{X_a : a \in \Sigma\} \cup \{Y_a : a \in \Sigma\}$ be a collection of independent and identically distributed standard normal random variables, and let $\mathcal{X} = \mathbb{C}^{\Sigma}$. Define a vector-valued random variable Z, taking values in $\mathcal{X}$, as

$$Z = \sum_{a \in \Sigma} (X_a + iY_a)e_a. \tag{7.118}$$

[1] Indeed, the measure μ is uniquely determined by these requirements. The fact that this is so will be verified through the use of the Haar measure, which is introduced below.

The *uniform spherical measure* μ on $\mathcal{S}(\mathcal{X})$ is the Borel probability measure

$$\mu : \text{Borel}(\mathcal{S}(\mathcal{X})) \to [0, 1] \tag{7.119}$$

defined as

$$\mu(\mathcal{A}) = \Pr\big(\alpha Z \in \mathcal{A} \text{ for some } \alpha > 0\big), \tag{7.120}$$

for every $\mathcal{A} \in \text{Borel}(\mathcal{S}(\mathcal{X}))$.

The fact that the uniform spherical measure μ is indeed a well-defined Borel probability measure follows from three simple observations. First, one has that

$$\big\{x \in \mathcal{X} \,:\, \alpha x \in \mathcal{A} \text{ for some } \alpha > 0\big\} = \text{cone}(\mathcal{A}) \backslash \{0\} \tag{7.121}$$

is a Borel subset of $\mathcal{X}$ for every Borel subset $\mathcal{A}$ of $\mathcal{S}(\mathcal{X})$, which implies that μ is a well-defined function. Second, if $\mathcal{A}$ and $\mathcal{B}$ are disjoint Borel subsets of $\mathcal{S}(\mathcal{X})$, then $\text{cone}(\mathcal{A}) \backslash \{0\}$ and $\text{cone}(\mathcal{B}) \backslash \{0\}$ are also disjoint, from which it follows that μ is a measure. Finally, it holds that

$$\mu(\mathcal{S}(\mathcal{X})) = \Pr(Z \neq 0) = 1, \tag{7.122}$$

and therefore μ is a probability measure.

It is evident that this definition is independent of how one might choose to order the elements of the alphabet Σ. For this reason, the fundamentally interesting properties of the uniform spherical measure defined on $\mathcal{S}(\mathcal{X})$ will follow from the same properties of the uniform spherical measure on $\mathcal{S}(\mathbb{C}^n)$. In some cases, restricting one's attention to complex Euclidean spaces of the form $\mathbb{C}^n$ will offer conveniences, mostly concerning notational simplicity, that will therefore cause no loss of generality.

The unitary invariance of the uniform spherical measure follows directly from the rotational invariance of the standard Gaussian measure, as the proof of the following proposition reveals.

Proposition 7.18. *For every complex Euclidean space $\mathcal{X}$, the uniform spherical measure μ on $\mathcal{S}(\mathcal{X})$ is unitarily invariant:*

$$\mu(U\mathcal{A}) = \mu(\mathcal{A}) \tag{7.123}$$

for every $\mathcal{A} \in \text{Borel}(\mathcal{S}(\mathcal{X}))$ and $U \in \text{U}(\mathcal{X})$.

Proof. Assume that Σ is the alphabet for which $\mathcal{X} = \mathbb{C}^{\Sigma}$, and let

$$\{X_a : a \in \Sigma\} \cup \{Y_a : a \in \Sigma\} \tag{7.124}$$

be a collection of independent and identically distributed standard normal random variables. Define vector-valued random variables X and Y, taking values in $\mathbb{R}^{\Sigma}$, as

$$X = \sum_{a \in \Sigma} X_a e_a \quad \text{and} \quad Y = \sum_{a \in \Sigma} Y_a e_a, \tag{7.125}$$

so that the vector-valued random variable Z referred to in Definition 7.17 may be expressed as $Z = X + iY$. To prove the proposition, it suffices to observe that Z and UZ are identically distributed for every unitary operator $U \in \mathrm{U}(\mathcal{X})$, for then one has that

$$\begin{aligned} \mu\big(U^{-1}\mathcal{A}\big) &= \Pr\big(\alpha UZ \in \mathcal{A} \text{ for some } \alpha > 0\big) \\ &= \Pr\big(\alpha Z \in \mathcal{A} \text{ for some } \alpha > 0\big) = \mu\big(\mathcal{A}\big) \end{aligned} \tag{7.126}$$

for every Borel subset $\mathcal{A}$ of $\mathcal{S}(\mathcal{X})$.

To verify that Z and UZ are identically distributed, for any choice of a unitary operator $U \in \mathrm{U}(\mathcal{X})$, note that

$$\begin{aligned} \begin{pmatrix} \Re(UZ) \\ \Im(UZ) \end{pmatrix} &= \begin{pmatrix} \Re(U) & -\Im(U) \\ \Im(U) & \Re(U) \end{pmatrix} \begin{pmatrix} \Re(Z) \\ \Im(Z) \end{pmatrix} \\ &= \begin{pmatrix} \Re(U) & -\Im(U) \\ \Im(U) & \Re(U) \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}, \end{aligned} \tag{7.127}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ denote the entry-wise real and imaginary parts of operators and vectors, as a calculation reveals. The operator

$$\begin{pmatrix} \Re(U) & -\Im(U) \\ \Im(U) & \Re(U) \end{pmatrix} \tag{7.128}$$

is an orthogonal operator on the vector space $\mathbb{R}^{\Sigma} \oplus \mathbb{R}^{\Sigma}$, while the vector-valued random variable $X \oplus Y$ is distributed with respect to the standard Gaussian measure on $\mathbb{R}^{\Sigma} \oplus \mathbb{R}^{\Sigma}$, and is therefore invariant under orthogonal transformations. It therefore follows that $X \oplus Y$ and $\Re(UZ) \oplus \Im(UZ)$ are identically distributed, which implies that Z and UZ are also identically distributed. □

Haar measure

Along similar lines to the uniform spherical measure, a unitarily invariant Borel probability measure η, known as the *Haar measure*,[2] may be defined on the set of unitary operators $\mathrm{U}(\mathcal{X})$ acting on given complex Euclidean space $\mathcal{X}$. More specifically, this measure is invariant with respect to both left and right multiplication by every unitary operator:

$$\eta(U\mathcal{A}) = \eta(\mathcal{A}) = \eta(\mathcal{A}U) \tag{7.129}$$

for every choice of $\mathcal{A} \in \mathrm{Borel}(\mathrm{U}(\mathcal{X}))$ and $U \in \mathrm{U}(\mathcal{X})$.

Definition 7.19. Let $\mathcal{X}$ be a complex Euclidean space, let Σ be the alphabet for which $\mathcal{X} = \mathbb{C}^{\Sigma}$, and let

$$\{X_{a,b} \,:\, a,b \in \Sigma\} \cup \{Y_{a,b} \,:\, a,b \in \Sigma\} \tag{7.130}$$

be a collection of independent and identically distributed standard normal random variables. Define an operator-valued random variable Z, taking values in $\mathrm{L}(\mathcal{X})$, as

$$Z = \sum_{a,b \in \Sigma} (X_{a,b} + iY_{a,b}) E_{a,b}. \tag{7.131}$$

The *Haar measure* η on $\mathrm{U}(\mathcal{X})$ is the Borel probability measure

$$\eta : \mathrm{Borel}(\mathrm{U}(\mathcal{X})) \to [0,1] \tag{7.132}$$

defined as

$$\eta(\mathcal{A}) = \Pr\big(PZ \in \mathcal{A} \text{ for some } P \in \mathrm{Pd}(\mathcal{X})\big) \tag{7.133}$$

for every $\mathcal{A} \in \mathrm{Borel}(\mathrm{U}(\mathcal{X}))$.

As the following theorem states, the Haar measure, as just defined, is indeed a Borel probability measure.

Theorem 7.20. *Let $\eta : \mathrm{Borel}(\mathrm{U}(\mathcal{X})) \to [0,1]$ be as specified by Definition 7.19, for any choice of a complex Euclidean space $\mathcal{X}$. It holds that η is a Borel probability measure.*

[2] The term *Haar measure* often refers to a more general notion, which is that of a measure defined on a certain class of groups that is invariant under the action of the group on which it is defined. The definition presented here is a restriction of this notion to the group of unitary operators acting on a given complex Euclidean space.

Proof. For every Borel subset $\mathcal{A} \subseteq \mathrm{U}(\mathcal{X})$, define a set $\mathcal{R}(\mathcal{A}) \subseteq \mathrm{L}(\mathcal{X})$ as

$$\mathcal{R}(\mathcal{A}) = \big\{QU \,:\, Q \in \mathrm{Pd}(\mathcal{X}) \text{ and } U \in \mathcal{A}\big\}. \tag{7.134}$$

For any operator $X \in \mathrm{L}(\mathcal{X})$, one has that $PX \in \mathcal{A}$ for some $P \in \mathrm{Pd}(\mathcal{X})$ if and only if $X \in \mathcal{R}(\mathcal{A})$. To prove that η is a Borel measure, it therefore suffices to prove that $\mathcal{R}(\mathcal{A})$ is a Borel subset of $\mathrm{L}(\mathcal{X})$ for every $\mathcal{A} \in \mathrm{Borel}(\mathrm{U}(\mathcal{X}))$, and that $\mathcal{R}(\mathcal{A})$ and $\mathcal{R}(\mathcal{B})$ are disjoint provided that $\mathcal{A}$ and $\mathcal{B}$ are disjoint.

The first of these requirements follows from the observation that the set $\mathrm{Pd}(\mathcal{X}) \times \mathcal{A}$ is a Borel subset of $\mathrm{Pd}(\mathcal{X}) \times \mathrm{U}(\mathcal{X})$, with respect to the product topology on the Cartesian product of these sets, together with the fact that operator multiplication is a continuous mapping.

For the second requirement, one observes that if

$$Q_0 U_0 = Q_1 U_1 \tag{7.135}$$

for some choice of $Q_0, Q_1 \in \mathrm{Pd}(\mathcal{X})$ and $U_0, U_1 \in \mathrm{U}(\mathcal{X})$, then it must hold that $Q_0 = Q_1 V$ for V being unitary. Therefore

$$Q_0^2 = Q_1 V V^* Q_1 = Q_1^2, \tag{7.136}$$

which implies that $Q_0 = Q_1$ by the fact that positive semidefinite operators have unique square roots. It therefore holds that $U_0 = U_1$. Consequently, if $\mathcal{R}(\mathcal{A}) \cap \mathcal{R}(\mathcal{B})$ is nonempty, then the same is true of $\mathcal{A} \cap \mathcal{B}$.

It remains to prove that η is a probability measure. Assume that Σ is the alphabet for which $\mathcal{X} = \mathbb{C}^{\Sigma}$, let

$$\{X_{a,b} \,:\, a, b \in \Sigma\} \cup \{Y_{a,b} \,:\, a, b \in \Sigma\} \tag{7.137}$$

be a collection of independent and identically distributed standard normal random variables, and define an operator-valued random variable

$$Z = \sum_{a,b \in \Sigma} (X_{a,b} + iY_{a,b}) E_{a,b}\,, \tag{7.138}$$

as in Definition 7.19. It holds that $PZ \in \mathrm{U}(\mathcal{X})$ for some positive definite operator $P \in \mathrm{Pd}(\mathcal{X})$ if and only if Z is nonsingular, and therefore

$$\eta(\mathrm{U}(\mathcal{X})) = \Pr\big(\mathrm{Det}(Z) \neq 0\big). \tag{7.139}$$

An operator is singular if and only if its column vectors form a linearly dependent set, and therefore $\mathrm{Det}(Z) = 0$ if and only if there exists a symbol $b \in \Sigma$ such that

$$\sum_{a\in\Sigma}(X_{a,b} + iY_{a,b})e_a \in \mathrm{span}\left\{\sum_{a\in\Sigma}(X_{a,c} + iY_{a,c})e_a \;:\; c \in \Sigma\backslash\{b\}\right\}. \tag{7.140}$$

The subspace referred to in this equation is necessarily a proper subspace of $\mathcal{X}$, because its dimension is at most $|\Sigma| - 1$, and therefore the event (7.140) occurs with probability zero. By the union bound, one has that

$$\Pr\big(\mathrm{Det}(Z) = 0\big) = 0, \tag{7.141}$$

as is implied by Proposition 1.21, and therefore $\eta(\mathrm{U}(\mathcal{X})) = 1$. □

The following proposition establishes that the Haar measure is unitarily invariant, in the sense specified by (7.129).

Proposition 7.21. *Let $\mathcal{X}$ be a complex Euclidean space. The Haar measure η on $\mathrm{U}(\mathcal{X})$ satisfies*

$$\eta(U\mathcal{A}) = \eta(\mathcal{A}) = \eta(\mathcal{A}U) \tag{7.142}$$

for every $\mathcal{A} \in \mathrm{Borel}(\mathrm{U}(\mathcal{X}))$ and $U \in \mathrm{U}(\mathcal{X})$.

Proof. Assume that Σ is the alphabet for which $\mathcal{X} = \mathbb{C}^\Sigma$, let

$$\{X_{a,b} \;:\; a, b \in \Sigma\} \cup \{Y_{a,b} \;:\; a, b \in \Sigma\} \tag{7.143}$$

be a collection of independent and identically distributed standard normal random variables, and let

$$Z = \sum_{a,b\in\Sigma}(X_{a,b} + iY_{a,b})E_{a,b}, \tag{7.144}$$

as in Definition 7.19.

Suppose that $\mathcal{A}$ is a Borel subset of $\mathrm{U}(\mathcal{X})$ and $U \in \mathrm{U}(\mathcal{X})$ is any unitary operator. To prove the left unitary invariance of η, it suffices to prove that Z and UZ are identically distributed, and to prove the right unitary invariance of η, it suffices to prove that Z and ZU are identically distributed, for then one has

$$\begin{aligned}\eta(U\mathcal{A}) &= \Pr\big(U^{-1}PZ \in \mathcal{A} \text{ for some } P \in \mathrm{Pd}(\mathcal{X})\big)\\ &= \Pr\big((U^{-1}PU)Z \in \mathcal{A} \text{ for some } P \in \mathrm{Pd}(\mathcal{X})\big) = \eta(\mathcal{A})\end{aligned} \tag{7.145}$$

and

$$\begin{aligned}\eta(\mathcal{A}U) &= \Pr\big(PZU^{-1} \in \mathcal{A} \text{ for some } P \in \mathrm{Pd}(\mathcal{X})\big) \\ &= \Pr\big(PZ \in \mathcal{A} \text{ for some } P \in \mathrm{Pd}(\mathcal{X})\big) = \eta(\mathcal{A}).\end{aligned} \tag{7.146}$$

The fact that UZ, Z, and ZU are identically distributed follows, through essentially the same argument as the one used to prove Proposition 7.18, from the invariance of the standard Gaussian measure under orthogonal transformations. □

For every complex Euclidean space, one has that the Haar measure η on $\mathrm{U}(\mathcal{X})$ is the unique Borel probability measure that is both left and right unitarily invariant. Indeed, any Borel probability measure on $\mathrm{U}(\mathcal{X})$ that is either left unitary invariant or right unitary invariant must necessarily be equal to the Haar measure, as the following theorem reveals.

Theorem 7.22. *Let $\mathcal{X}$ be a complex Euclidean space and let*

$$\nu : \mathrm{Borel}(\mathrm{U}(\mathcal{X})) \to [0,1] \tag{7.147}$$

be a Borel probability measure possessing either of the following two properties:

1. *Left unitary invariance: $\nu(U\mathcal{A}) = \nu(\mathcal{A})$ for every Borel subset $\mathcal{A} \subseteq \mathrm{U}(\mathcal{X})$ and every unitary operator $U \in \mathrm{U}(\mathcal{X})$.*
2. *Right unitary invariance: $\nu(\mathcal{A}U) = \nu(\mathcal{A})$ for every Borel subset $\mathcal{A} \subseteq \mathrm{U}(\mathcal{X})$ and every unitary operator $U \in \mathrm{U}(\mathcal{X})$.*

It holds that ν is equal to the Haar measure $\eta : \mathrm{Borel}(\mathrm{U}(\mathcal{X})) \to [0,1]$.

Proof. It will be assumed that ν is left unitarily invariant; the case in which ν is right unitarily invariant is proved through a similar argument. Let $\mathcal{A}$ be an arbitrary Borel subset of $\mathrm{U}(\mathcal{X})$, and let f denote the characteristic function of $\mathcal{A}$:

$$f(U) = \begin{cases} 1 & \text{if } U \in \mathcal{A} \\ 0 & \text{if } U \notin \mathcal{A} \end{cases} \tag{7.148}$$

for every $U \in \mathrm{U}(\mathcal{X})$. One has that

$$\nu(\mathcal{A}) = \int f(U)\,\mathrm{d}\nu(U) = \int f(VU)\,\mathrm{d}\nu(U) \tag{7.149}$$

for every unitary operator $V \in \mathrm{U}(\mathcal{X})$ by the left unitary invariance of ν. Integrating over all unitary operators V with respect to the Haar measure η yields

$$\nu(\mathcal{A}) = \iint f(VU)\,\mathrm{d}\nu(U)\,\mathrm{d}\eta(V) = \iint f(VU)\,\mathrm{d}\eta(V)\,\mathrm{d}\nu(U), \tag{7.150}$$

where the change in the order of integration is made possible by Fubini's theorem. By the right unitary invariance of Haar measure, it follows that

$$\nu(\mathcal{A}) = \iint f(V)\,\mathrm{d}\eta(V)\,\mathrm{d}\nu(U) = \int f(V)\,\mathrm{d}\eta(V) = \eta(\mathcal{A}). \tag{7.151}$$

As $\mathcal{A}$ was chosen arbitrarily, it follows that $\nu = \eta$, as required. □

The Haar measure and uniform spherical measure are closely related, as the following theorem indicates. The proof uses the same methodology as the proof of the previous theorem.

Theorem 7.23. *Let $\mathcal{X}$ be a complex Euclidean space, let μ denote the uniform spherical measure on $\mathcal{S}(\mathcal{X})$, and let η denote the Haar measure on $\mathrm{U}(\mathcal{X})$. For every $\mathcal{A} \in \mathrm{Borel}(\mathcal{S}(\mathcal{X}))$ and $x \in \mathcal{S}(\mathcal{X})$, it holds that*

$$\mu(\mathcal{A}) = \eta\big(\{U \in \mathrm{U}(\mathcal{X}) \,:\, Ux \in \mathcal{A}\}\big). \tag{7.152}$$

Proof. Let $\mathcal{A}$ be any Borel subset of $\mathcal{S}(\mathcal{X})$ and let f denote the characteristic function of f:

$$f(y) = \begin{cases} 1 & \text{if } y \in \mathcal{A} \\ 0 & \text{if } y \notin \mathcal{A} \end{cases} \tag{7.153}$$

for every $y \in \mathcal{S}(\mathcal{X})$. It holds that

$$\mu(\mathcal{A}) = \int f(y)\,\mathrm{d}\mu(y) = \int f(Uy)\,\mathrm{d}\mu(y) \tag{7.154}$$

for every $U \in \mathrm{U}(\mathcal{X})$, by the unitary invariance of the uniform spherical measure. Integrating over all $U \in \mathrm{U}(\mathcal{X})$ with respect to the Haar measure and changing the order of integration by means of Fubini's theorem yields

$$\mu(\mathcal{A}) = \iint f(Uy)\,\mathrm{d}\mu(y)\,\mathrm{d}\eta(U) = \iint f(Uy)\,\mathrm{d}\eta(U)\,\mathrm{d}\mu(y). \tag{7.155}$$

Now, for any fixed choice of unit vectors $x, y \in \mathcal{S}(\mathcal{X})$, one may choose a unitary operator $V_{x,y} \in \mathrm{U}(\mathcal{X})$ for which it holds that $V_{x,y}x = y$. By the right unitary invariance of the Haar measure, one has

$$\int f(Uy)\,\mathrm{d}\eta(U) = \int f(UV_{x,y}y)\,\mathrm{d}\eta(U) = \int f(Ux)\,\mathrm{d}\eta(U). \tag{7.156}$$

Consequently,

$$\begin{aligned}\mu(\mathcal{A}) &= \iint f(Uy)\,\mathrm{d}\eta(U)\,\mathrm{d}\mu(y) = \iint f(Ux)\,\mathrm{d}\eta(U)\,\mathrm{d}\mu(y) \\ &= \int f(Ux)\,\mathrm{d}\eta(U) = \eta\big(\{U \in \mathrm{U}(\mathcal{X}) \,:\, Ux \in \mathcal{A}\}\big),\end{aligned} \tag{7.157}$$

as required. □

Noting that the proof of the previous theorem has not made use of any properties of the measure μ aside from the fact that it is normalized and unitarily invariant, one obtains the following corollary.

Corollary 7.24. *Let $\mathcal{X}$ be a complex Euclidean space and let*

$$\nu : \mathrm{Borel}(\mathcal{S}(\mathcal{X})) \to [0,1] \tag{7.158}$$

be a Borel probability measure that is unitarily invariant: $\nu(U\mathcal{A}) = \nu(\mathcal{A})$ for every Borel subset $\mathcal{A} \subseteq \mathcal{S}(\mathcal{X})$. It holds that ν is equal to the uniform spherical measure $\mu : \mathrm{Borel}(\mathcal{S}(\mathcal{X})) \to [0,1]$.

Evaluating integrals by means of symmetries

Some integrals defined with respect to the uniform spherical measure or Haar measure may be evaluated by considering the symmetries present in those integrals. For example, for Σ being any alphabet and μ denoting the uniform spherical measure on $\mathcal{S}\big(\mathbb{C}^{\Sigma}\big)$, one has that

$$\int uu^*\mathrm{d}\mu(u) = \frac{\mathbb{1}}{|\Sigma|}. \tag{7.159}$$

This is so because the operator represented by the integral is necessarily positive semidefinite, has unit trace, and is invariant under conjugation by every unitary operator—and the only operator having these properties is $\mathbb{1}/|\Sigma|$.

The following lemma establishes a generalization of this fact, providing an alternative description of the projection onto the symmetric subspace defined in Section 7.1.1.

Lemma 7.25. *Let n be a positive integer, let Σ be an alphabet, let $\mathcal{X}_k = \mathbb{C}^\Sigma$ for $k = 1, \ldots, n$, and let μ denote the uniform spherical measure on $\mathcal{S}\big(\mathbb{C}^\Sigma\big)$. It holds that*

$$\Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n} = \binom{|\Sigma| + n - 1}{|\Sigma| - 1} \int (uu^*)^{\otimes n} \mathrm{d}\mu(u). \tag{7.160}$$

Proof. Let

$$P = \binom{|\Sigma| + n - 1}{|\Sigma| - 1} \int (uu^*)^{\otimes n} \, \mathrm{d}\mu(u), \tag{7.161}$$

and note first that

$$\mathrm{Tr}(P) = \binom{|\Sigma| + n - 1}{|\Sigma| - 1} = \mathrm{Tr}\big(\Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}\big), \tag{7.162}$$

as μ is a normalized measure.

Next, by the unitary invariance of the uniform spherical measure, one has that $[P, U^{\otimes n}] = 0$ for every $U \in \mathrm{U}(\mathbb{C}^\Sigma)$. By Theorem 7.16, it follows that

$$P = \sum_{\pi \in S_n} u(\pi) W_\pi \tag{7.163}$$

for some choice of a vector $u \in \mathbb{C}^{S_n}$. Using the fact that

$$u^{\otimes n} \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \tag{7.164}$$

for every unit vector $u \in \mathbb{C}^\Sigma$, one necessarily has that

$$\Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n} P = P, \tag{7.165}$$

which implies

$$\begin{aligned} P &= \frac{1}{n!} \sum_{\sigma \in S_n} W_\sigma \sum_{\pi \in S_n} u(\pi) W_\pi = \frac{1}{n!} \sum_{\pi \in S_n} \sum_{\sigma \in S_n} u(\sigma^{-1}\pi) W_\pi \\ &= \frac{1}{n!} \sum_{\sigma \in S_n} u(\sigma) \sum_{\pi \in S_n} W_\pi = \sum_{\sigma \in S_n} u(\sigma) \Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n} \end{aligned} \tag{7.166}$$

by Proposition 7.1. By (7.162), one has

$$\sum_{\sigma \in S_n} u(\sigma) = 1, \tag{7.167}$$

and therefore $P = \Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}$, as required. □

The following example represents a continuation of Example 6.11. Two channels that have a close connection to the classes of Werner states and isotropic states are analyzed based on properties of their symmetries.

Example 7.26. As in Example 6.11, let Σ be an alphabet, let $n = |\Sigma|$, and let $\mathcal{X} = \mathbb{C}^{\Sigma}$, and recall the four projection operators[3]

$$\Delta_0,\, \Delta_1,\, \Pi_0,\, \Pi_1 \in \operatorname{Proj}(\mathcal{X} \otimes \mathcal{X}) \tag{7.168}$$

defined in that example:

$$\Delta_0 = \frac{1}{n} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b}, \tag{7.169}$$

$$\Delta_1 = \mathbb{1} \otimes \mathbb{1} - \frac{1}{n} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b}\,, \tag{7.170}$$

$$\Pi_0 = \frac{1}{2}\mathbb{1} \otimes \mathbb{1} + \frac{1}{2} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b}\,, \tag{7.171}$$

$$\Pi_1 = \frac{1}{2}\mathbb{1} \otimes \mathbb{1} - \frac{1}{2} \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{a,b}\,. \tag{7.172}$$

Equivalently, one may write

$$\Delta_0 = \frac{1}{n}(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(W)\,, \qquad \Pi_0 = \frac{1}{2}\mathbb{1} \otimes \mathbb{1} + \frac{1}{2}W\,, \tag{7.173}$$

$$\Delta_1 = \mathbb{1} \otimes \mathbb{1} - \frac{1}{n}(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(W)\,, \quad \Pi_1 = \frac{1}{2}\mathbb{1} \otimes \mathbb{1} - \frac{1}{2}W\,, \tag{7.174}$$

for $\mathrm{T}(X) = X^{\mathsf{T}}$ denoting the transpose mapping on $\mathrm{L}(\mathcal{X})$ and

$$W = \sum_{a,b \in \Sigma} E_{a,b} \otimes E_{b,a}\,, \tag{7.175}$$

which is the swap operator on $\mathcal{X} \otimes \mathcal{X}$. States of the form

$$\lambda \Delta_0 + (1 - \lambda)\frac{\Delta_1}{n^2 - 1} \quad \text{and} \quad \lambda \frac{\Pi_0}{\binom{n+1}{2}} + (1 - \lambda)\frac{\Pi_1}{\binom{n}{2}}\,, \tag{7.176}$$

[3] Using the notation introduced in Section 7.1.1, one may alternately write $\Pi_0 = \Pi_{\mathcal{X} \otimes \mathcal{X}}$ and $\Pi_1 = \Pi_{\mathcal{X} \oslash \mathcal{X}}$. The notations Π_0 and Π_1 will be used within this example to maintain consistency with Example 6.11.

for $\lambda \in [0,1]$, were introduced in Example 6.11 as *isotropic states* and *Werner states*, respectively.

Now, consider the channel $\Xi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{X})$ defined as

$$\Xi(X) = \int (U \otimes U) X (U \otimes U)^* \, \mathrm{d}\eta(U) \tag{7.177}$$

for all $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$, for η denoting the Haar measure on $\mathrm{U}(\mathcal{X})$. By the unitary invariance of Haar measure, one has that $[\Xi(X), U \otimes U] = 0$ for every $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$ and $U \in \mathrm{U}(\mathcal{X})$. By Theorem 7.16 it holds that

$$\Xi(X) \in \operatorname{span}\{\mathbb{1} \otimes \mathbb{1}, W\} = \operatorname{span}\{\Pi_0, \Pi_1\}, \tag{7.178}$$

and it must therefore hold that

$$\Xi(X) = \alpha(X)\,\Pi_0 + \beta(X)\,\Pi_1 \tag{7.179}$$

for $\alpha(X), \beta(X) \in \mathbb{C}$ being complex numbers depending linearly on X. The channel Ξ is self-adjoint and satisfies $\Xi(\mathbb{1} \otimes \mathbb{1}) = \mathbb{1} \otimes \mathbb{1}$ and $\Xi(W) = W$, so that $\Xi(\Pi_0) = \Pi_0$ and $\Xi(\Pi_1) = \Pi_1$. The following two equations hold:

$$\begin{aligned} \alpha(X) &= \frac{1}{\binom{n+1}{2}} \langle \Pi_0, \Xi(X) \rangle = \frac{1}{\binom{n+1}{2}} \langle \Xi(\Pi_0), X \rangle = \frac{1}{\binom{n+1}{2}} \langle \Pi_0, X \rangle \\ \beta(X) &= \frac{1}{\binom{n}{2}} \langle \Pi_1, \Xi(X) \rangle = \frac{1}{\binom{n}{2}} \langle \Xi(\Pi_1), X \rangle = \frac{1}{\binom{n}{2}} \langle \Pi_1, X \rangle. \end{aligned} \tag{7.180}$$

It therefore follows that

$$\Xi(X) = \frac{1}{\binom{n+1}{2}} \langle \Pi_0, X \rangle \Pi_0 + \frac{1}{\binom{n}{2}} \langle \Pi_1, X \rangle \Pi_1. \tag{7.181}$$

It is evident from this expression that, on any density operator input, the output of the channel Ξ is a Werner state, and moreover every Werner state is fixed by this channel. The channel Ξ is sometimes called a *Werner twirling channel*.

A different but closely related channel $\Lambda \in \mathrm{C}(\mathcal{X} \otimes \mathcal{X})$ is defined as

$$\Lambda(X) = \int (U \otimes \overline{U}) X (U \otimes \overline{U})^* \, \mathrm{d}\eta(U) \tag{7.182}$$

for all $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$, where η again denotes the Haar measure on $\mathrm{U}(\mathcal{X})$. An alternate expression of this channel may be obtained by making use of

the analysis of the channel Ξ presented above. The first step of this process is to observe that Λ may be obtained by composing the channel Ξ with the partial transpose in the following way:

$$\Lambda = (\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})\, \Xi\, (\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}). \tag{7.183}$$

Then, using the identities

$$\begin{aligned} (\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(\Pi_0) &= \frac{n+1}{2}\Delta_0 + \frac{1}{2}\Delta_1, \\ (\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(\Pi_1) &= -\frac{n-1}{2}\Delta_0 + \frac{1}{2}\Delta_1, \end{aligned} \tag{7.184}$$

one finds that

$$\Lambda(X) = \langle \Delta_0, X \rangle \Delta_0 + \frac{1}{n^2-1} \langle \Delta_1, X \rangle \Delta_1. \tag{7.185}$$

On any density operator input, the output of the channel Λ is an isotropic state, and moreover every isotropic state is fixed by Λ. The channel Λ is sometimes called an *isotropic twirling channel*.

It is evident from the specification of the channels Ξ and Λ that one has the following expressions, in which Φ_U denotes the unitary channel defined by $\Phi_U(X) = UXU^*$ for each $X \in \mathrm{L}(\mathcal{X})$:

$$\begin{aligned} \Xi &\in \operatorname{conv}\bigl\{\Phi_U \otimes \Phi_U \,:\, U \in \mathrm{U}(\mathcal{X})\bigr\}, \\ \Lambda &\in \operatorname{conv}\bigl\{\Phi_U \otimes \Phi_{\overline{U}} \,:\, U \in \mathrm{U}(\mathcal{X})\bigr\}. \end{aligned} \tag{7.186}$$

It follows that Ξ and Λ are mixed-unitary channels, and LOCC channels as well. Indeed, both channels can be implemented without communication—local operations and shared randomness are sufficient.

Finally, for any choice of orthogonal unit vectors $x, y \in \mathcal{X}$, the following equalities may be observed:

$$\begin{aligned} \bigl\langle \Pi_0, xx^* \otimes yy^* \bigr\rangle &= \frac{1}{2}, \quad \bigl\langle \Pi_1, xx^* \otimes yy^* \bigr\rangle = \frac{1}{2}, \\ \bigl\langle \Pi_0, xx^* \otimes xx^* \bigr\rangle &= 1, \quad \bigl\langle \Pi_1, xx^* \otimes xx^* \bigr\rangle = 0. \end{aligned} \tag{7.187}$$

Therefore, for every choice of $\alpha \in [0,1]$, one has

$$\Xi(xx^* \otimes (\alpha xx^* + (1-\alpha)yy^*)) = \Bigl(1 - \frac{\alpha}{2}\Bigr) \frac{\Pi_0}{\binom{n+1}{2}} + \frac{\alpha}{2} \frac{\Pi_1}{\binom{n}{2}}. \tag{7.188}$$

As Ξ is a separable channel and

$$xx^* \otimes (\alpha xx^* + (1-\alpha)yy^*) \in \mathrm{SepD}(\mathcal{X} : \mathcal{X}) \tag{7.189}$$

is a separable state, for every $\alpha \in [0,1]$, it follows that the state (7.188) is also separable. Equivalently, the Werner state

$$\lambda \frac{\Pi_0}{\binom{n+1}{2}} + (1-\lambda)\frac{\Pi_1}{\binom{n}{2}} \tag{7.190}$$

is separable for all $\lambda \in [1/2, 1]$. The partial transpose of the state (7.190) is

$$\frac{2\lambda - 1}{n}\Delta_0 + \left(1 - \frac{2\lambda - 1}{n}\right)\frac{\Delta_1}{n^2 - 1}. \tag{7.191}$$

Under the assumption that $\lambda \in [1/2, 1]$, the state (7.190) is separable, and therefore its partial transpose is also separable. It follows that the isotropic state

$$\lambda\Delta_0 + (1-\lambda)\frac{\Delta_1}{n^2-1} \tag{7.192}$$

is separable for all $\lambda \in [0, 1/n]$.

7.2.2 Applications of unitarily invariant measures

Integration with respect to uniform spherical measure and Haar measure are useful tools in the theory of quantum information, and there exist many interesting examples of applications that rely upon them. Three examples are presented below, and further examples involving the phenomenon of *measure concentration* are presented in Section 7.3.2.

The quantum de Finetti theorem

Intuitively speaking, the quantum de Finetti theorem states that if the state of a collection of identical registers is exchangeable, then the reduced state of any comparatively small number of these registers must be close to a convex combination of identical product states. This theorem will first be stated and proved for symmetric pure states, and from this theorem a more general statement for arbitrary exchangeable states may be derived using Theorem 7.13.

Theorem 7.27. *Let n be a positive integer, let Σ be an alphabet, and let $\mathsf{X}_1, \ldots, \mathsf{X}_n$ be registers sharing the same classical state set Σ. For every symmetric unit vector $v \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$ and every positive integer $k \in \{1, \ldots, n\}$, there exists a density operator*

$$\tau \in \operatorname{conv}\Big\{ (uu^*)^{\otimes k} \,:\, u \in \mathcal{S}\big(\mathbb{C}^\Sigma\big) \Big\} \tag{7.193}$$

such that

$$\big\| (vv^*)[\mathsf{X}_1, \ldots, \mathsf{X}_k] - \tau \big\|_1 \leq \frac{4k\big(|\Sigma| - 1\big)}{n+1}. \tag{7.194}$$

Proof. It will be proved that the requirements of the theorem are satisfied by the operator

$$\tau = \binom{n + |\Sigma| - 1}{|\Sigma| - 1} \int \langle (uu^*)^{\otimes n}, vv^* \rangle (uu^*)^{\otimes k} \, \mathrm{d}\mu(u), \tag{7.195}$$

for μ denoting the uniform spherical measure on $\mathcal{S}\big(\mathbb{C}^\Sigma\big)$. The fact that τ is positive semidefinite is evident from its definition, and by Lemma 7.25, together with the assumption $v \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, one has that $\operatorname{Tr}(\tau) = 1$.

For the sake of establishing the bound (7.194), it is convenient to define

$$N_m = \binom{m + |\Sigma| - 1}{|\Sigma| - 1} \tag{7.196}$$

for every nonnegative integer m. The following bounds on the ratio between N_{n-k} and N_n hold:

$$\begin{aligned} 1 \geq \frac{N_{n-k}}{N_n} &= \frac{n - k + |\Sigma| - 1}{n + |\Sigma| - 1} \cdots \frac{n - k + 1}{n + 1} \\ &\geq \left(\frac{n - k + 1}{n + 1} \right)^{|\Sigma| - 1} \geq 1 - \frac{k\big(|\Sigma| - 1\big)}{n + 1}. \end{aligned} \tag{7.197}$$

For every unit vector $u \in \mathcal{S}(\mathbb{C}^\Sigma)$ and every positive integer m, define a projection operator

$$\Delta_{m,u} = (uu^*)^{\otimes m}, \tag{7.198}$$

and also define an operator $P_u \in \operatorname{Pos}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k)$ as

$$P_u = \operatorname{Tr}_{\mathcal{X}_{k+1} \otimes \cdots \otimes \mathcal{X}_n} \Big(\big(\mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \Delta_{n-k,u} \big) vv^* \Big). \tag{7.199}$$

By Lemma 7.25, together with the assumption $v \in \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n$, one has that

$$vv^* = N_{n-k} \int \left(\mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_k} \otimes \Delta_{n-k,u}\right) vv^* \mathrm{d}\mu(u), \tag{7.200}$$

and therefore

$$\left(vv^*\right)[\mathsf{X}_1, \ldots, \mathsf{X}_k] = N_{n-k} \int P_u \, \mathrm{d}\mu(u). \tag{7.201}$$

This density operator is to be compared with τ, which may be expressed as

$$\tau = N_n \int \Delta_{k,u} P_u \Delta_{k,u} \, \mathrm{d}\mu(u). \tag{7.202}$$

The primary goal of the remainder of the proof is to bound the trace norm of the operator

$$\frac{1}{N_{n-k}} \left(vv^*\right)[\mathsf{X}_1, \ldots, \mathsf{X}_k] - \frac{1}{N_n} \tau = \int \Big(P_u - \Delta_{k,u} P_u \Delta_{k,u} \Big) \, \mathrm{d}\mu(u), \tag{7.203}$$

as such a bound will lead directly to a bound on the trace norm of

$$\left(vv^*\right)[\mathsf{X}_1, \ldots, \mathsf{X}_k] - \tau. \tag{7.204}$$

The operator identity

$$A - BAB = A(\mathbb{1} - B) + (\mathbb{1} - B)A - (\mathbb{1} - B)A(\mathbb{1} - B), \tag{7.205}$$

which holds for any two square operators A and B acting on a given space, will be useful for this purpose. It holds that

$$\int P_u (\mathbb{1} - \Delta_{k,u}) \, \mathrm{d}\mu(u) = \left(\frac{1}{N_{n-k}} - \frac{1}{N_n} \right) \left(vv^*\right)[\mathsf{X}_1, \ldots, \mathsf{X}_k] \tag{7.206}$$

and therefore

$$\left\| \int P_u (\mathbb{1} - \Delta_{k,u}) \, \mathrm{d}\mu(u) \right\|_1 = \left(\frac{1}{N_{n-k}} - \frac{1}{N_n} \right). \tag{7.207}$$

By similar reasoning, one finds that

$$\left\| \int (\mathbb{1} - \Delta_{k,u}) P_u \, \mathrm{d}\mu(u) \right\|_1 = \left(\frac{1}{N_{n-k}} - \frac{1}{N_n} \right). \tag{7.208}$$

Moreover, one has

$$\begin{aligned}
&\left\| \int (\mathbb{1} - \Delta_{k,u}) P_u (\mathbb{1} - \Delta_{k,u}) \, \mathrm{d}\mu(u) \right\|_1 \\
&\qquad = \mathrm{Tr}\left(\int (\mathbb{1} - \Delta_{k,u}) P_u (\mathbb{1} - \Delta_{k,u}) \, \mathrm{d}\mu(u) \right) \\
&\qquad = \mathrm{Tr}\left(\int P_u (\mathbb{1} - \Delta_{k,u}) \, \mathrm{d}\mu(u) \right) = \left(\frac{1}{N_{n-k}} - \frac{1}{N_n} \right),
\end{aligned} \tag{7.209}$$

and therefore, by the triangle inequality together with the identity (7.205), it follows that

$$\left\| \frac{1}{N_{n-k}} (vv^*)[\mathsf{X}_1, \ldots, \mathsf{X}_k] - \frac{1}{N_n} \tau \right\|_1 \leq 3 \left(\frac{1}{N_{n-k}} - \frac{1}{N_n} \right). \tag{7.210}$$

Having established a bound on the trace norm of the operator (7.203), the theorem follows:

$$\begin{aligned}
&\left\| (vv^*)[\mathsf{X}_1, \ldots, \mathsf{X}_k] - \tau \right\|_1 \\
&\qquad \leq N_{n-k} \left\| \frac{1}{N_{n-k}} (vv^*)[\mathsf{X}_1, \ldots, \mathsf{X}_k] - \frac{1}{N_n} \tau \right\|_1 \\
&\qquad\qquad + N_{n-k} \left\| \frac{1}{N_n} \tau - \frac{1}{N_{n-k}} \tau \right\|_1 \\
&\qquad \leq 4 \left(1 - \frac{N_{n-k}}{N_n} \right) \\
&\qquad \leq \frac{4k\big(|\Sigma| - 1\big)}{n+1},
\end{aligned} \tag{7.211}$$

as required. □

Corollary 7.28 (Quantum de Finetti theorem)**.** *Let n be a positive integer, let Σ be an alphabet, and let $\mathsf{X}_1, \ldots, \mathsf{X}_n$ be registers sharing the same classical state set Σ. For every exchangeable density operator $\rho \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$ and every positive integer $k \in \{1, \ldots, n\}$, there exists a density operator*

$$\tau \in \mathrm{conv}\big\{ \sigma^{\otimes k} \,:\, \sigma \in \mathrm{D}\big(\mathbb{C}^{\Sigma}\big) \big\} \tag{7.212}$$

such that

$$\big\| \rho[\mathsf{X}_1, \ldots, \mathsf{X}_k] - \tau \big\|_1 \leq \frac{4k\big(|\Sigma|^2 - 1\big)}{n+1}. \tag{7.213}$$

Proof. Let $\mathsf{Y}_1, \ldots, \mathsf{Y}_n$ be registers, all sharing the classical state set Σ. By Theorem 7.13, there exists a symmetric unit vector

$$v \in (\mathcal{X}_1 \otimes \mathcal{Y}_1) \otimes \cdots \otimes (\mathcal{X}_n \otimes \mathcal{Y}_n), \tag{7.214}$$

representing a pure state of the compound register $((\mathsf{X}_1, \mathsf{Y}_1), \ldots, (\mathsf{X}_n, \mathsf{Y}_n))$, with the property that

$$(vv^*)[\mathsf{X}_1, \ldots, \mathsf{X}_n] = \rho. \tag{7.215}$$

By Theorem 7.27, there exists a density operator

$$\xi \in \operatorname{conv}\bigl\{(uu^*)^{\otimes k} \,:\, u \in \mathcal{S}\bigl(\mathbb{C}^\Sigma \otimes \mathbb{C}^\Sigma\bigr)\bigr\}, \tag{7.216}$$

representing a state of the compound register $((\mathsf{X}_1, \mathsf{Y}_1), \ldots, (\mathsf{X}_k, \mathsf{Y}_k))$, such that

$$\Bigl\| (vv^*)[(\mathsf{X}_1, \mathsf{Y}_1), \ldots, (\mathsf{X}_k, \mathsf{Y}_k)] - \xi \Bigr\|_1 \leq \frac{4k\bigl(|\Sigma|^2 - 1\bigr)}{n+1}. \tag{7.217}$$

Taking $\tau = \xi[\mathsf{X}_1, \ldots, \mathsf{X}_k]$, one has that

$$\tau \in \operatorname{conv}\bigl\{\sigma^{\otimes k} \,:\, \sigma \in \mathrm{D}\bigl(\mathbb{C}^\Sigma\bigr)\bigr\}, \tag{7.218}$$

and the required bound

$$\begin{aligned} \bigl\|\rho[\mathsf{X}_1, \ldots, \mathsf{X}_k] - \tau\bigr\|_1 &\leq \bigl\| (vv^*)[(\mathsf{X}_1, \mathsf{Y}_1), \ldots, (\mathsf{X}_k, \mathsf{Y}_k)] - \xi \bigr\|_1 \\ &\leq \frac{4k\bigl(|\Sigma|^2 - 1\bigr)}{n+1} \end{aligned} \tag{7.219}$$

follows by the monotonicity of the trace norm under partial tracing. □

Optimal cloning of pure quantum states

Let Σ be an alphabet, let n and m be positive integers with $n \leq m$, and let $\mathsf{X}_1, \ldots, \mathsf{X}_m$ be registers, all sharing the same classical state Σ. In the task of *cloning*, one assumes that the state of $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ is given by

$$\rho^{\otimes n} \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n), \tag{7.220}$$

for some choice of $\rho \in \mathrm{D}(\mathbb{C}^\Sigma)$, and the goal is to transform $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ into $(\mathsf{X}_1, \ldots, \mathsf{X}_m)$ in such a way that the resulting state of this register is as close as possible to

$$\rho^{\otimes m} \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m). \tag{7.221}$$

One may consider the quality with which a given channel

$$\Phi \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m) \tag{7.222}$$

performs this task in a variety of specific ways. For example, one might measure the closeness of $\Phi(\rho^n)$ to ρ^m with respect to the trace norm, some other norm, or the fidelity function; and one might consider the average closeness over some distribution on the possible choices of ρ, or consider the worst case over all ρ or over some subset of possible choices for ρ. It is most typical that one assumes ρ is a pure state—the mixed state case is more complicated and has very different characteristics from the pure state case.

The specific variant of the cloning task that will be considered here is that one aims to choose a channel of the form (7.222) so as to maximize the minimum fidelity

$$\alpha(\Phi) = \inf_{u \in \mathcal{S}(\mathbb{C}^\Sigma)} \mathrm{F}\big(\Phi\big((uu^*)^{\otimes n}\big), (uu^*)^{\otimes m}\big) \tag{7.223}$$

over all pure states $\rho = uu^*$. The following theorem establishes an upper bound on this quantity, and states that this bound is achieved for some choice of a channel Φ.

Theorem 7.29 (Werner). *Let Σ be an alphabet, let n and m be positive integers with $n \leq m$, and let $\mathcal{X}_1, \ldots, \mathcal{X}_m$ be complex Euclidean spaces with $\mathcal{X}_k = \mathbb{C}^\Sigma$ for all $k = 1, \ldots, m$. For every channel $\Phi \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m)$ it holds that*

$$\inf_{u \in \mathcal{S}(\mathbb{C}^\Sigma)} \big\langle \Phi\big((uu^*)^{\otimes n}\big), (uu^*)^{\otimes m} \big\rangle \leq \frac{N_n}{N_m}, \tag{7.224}$$

where

$$N_k = \binom{k + |\Sigma| - 1}{|\Sigma| - 1} \tag{7.225}$$

for each positive integer k. Moreover, there exists a channel Φ of the above form for which equality is achieved in (7.224).

Proof. The infimum in (7.224) can be no larger than the average with respect to the uniform spherical measure on $\mathcal{S}(\mathbb{C}^\Sigma)$:

$$\begin{aligned} &\inf_{u \in \mathcal{S}(\mathbb{C}^\Sigma)} \big\langle \Phi\big((uu^*)^{\otimes n}\big), (uu^*)^{\otimes m} \big\rangle \\ &\qquad \leq \int \big\langle \Phi\big((uu^*)^{\otimes n}\big), (uu^*)^{\otimes m} \big\rangle \, \mathrm{d}\mu(u). \end{aligned} \tag{7.226}$$

As $(uu^*)^{\otimes n} \leq \Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}$ for every $u \in \mathcal{S}(\mathbb{C}^\Sigma)$, it follows that

$$\begin{aligned}
&\int \big\langle \Phi\big((uu^*)^{\otimes n}\big), (uu^*)^{\otimes m} \big\rangle \, \mathrm{d}\mu(u) \\
&\qquad \leq \int \big\langle \Phi\big(\Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}\big), (uu^*)^{\otimes m} \big\rangle \, \mathrm{d}\mu(u) \\
&\qquad = \frac{1}{N_m} \big\langle \Phi\big(\Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}\big), \Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m} \big\rangle \\
&\qquad \leq \frac{1}{N_m} \operatorname{Tr}\big(\Phi\big(\Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}\big)\big) \\
&\qquad = \frac{N_n}{N_m}.
\end{aligned} \tag{7.227}$$

This establish the required bound (7.224).

It remains to prove that there exists a channel

$$\Phi \in \mathrm{C}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n, \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m) \tag{7.228}$$

for which equality is achieved in (7.224). Define

$$\begin{aligned}
\Psi(X) = \frac{N_n}{N_m} \Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m} \big(X \otimes \mathbb{1}_{\mathcal{X}_{n+1} \otimes \cdots \otimes \mathcal{X}_m}\big) \Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m} \\
+ \big\langle \mathbb{1}_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n} - \Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}, X \big\rangle \, \sigma
\end{aligned} \tag{7.229}$$

for all $X \in \mathrm{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n)$, where $\sigma \in \mathrm{D}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m)$ is an arbitrary density operator. It is evident that Φ is completely positive, and the fact that Φ is trace-preserving follows from the observation

$$\operatorname{Tr}_{\mathcal{X}_{n+1} \otimes \cdots \otimes \mathcal{X}_m} (\Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m}) = \frac{N_m}{N_n} \Pi_{\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n}. \tag{7.230}$$

A direct calculation reveals that

$$\big\langle (uu^*)^{\otimes m}, \Psi\big((uu^*)^{\otimes n}\big) \big\rangle = \frac{N_n}{N_m} \tag{7.231}$$

for every unit vector $u \in \mathbb{C}^\Sigma$, which completes the proof. □

Example 7.30. The channel described in Example 2.33 is an optimal cloning channel, achieving equality in (7.224) for the case $\Sigma = \{0,1\}$, $n = 1$, and $m = 2$.

Unital channels near the completely depolarizing channel

The final example of an application of unitarily invariant measures in the theory of quantum information to be presented in this section demonstrates that all unital channels sufficiently close to the completely depolarizing channel must be mixed-unitary channels. The following lemma will be used to demonstrate this fact.

Lemma 7.31. *Let $\mathcal{X}$ be a complex Euclidean space having dimension $n \geq 2$, let η denote the Haar measure on $\mathrm{U}(\mathcal{X})$, and let $\Omega \in \mathrm{C}(\mathcal{X})$ denote the completely depolarizing channel with respect to the space $\mathcal{X}$. The map $\Xi \in \mathrm{CP}(\mathcal{X} \otimes \mathcal{X})$ defined as*

$$\Xi(X) = \int \langle \operatorname{vec}(U) \operatorname{vec}(U)^*, X \rangle \operatorname{vec}(U) \operatorname{vec}(U)^* \, \mathrm{d}\eta(U) \tag{7.232}$$

for every $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$ is given by

$$\Xi = \frac{1}{n^2 - 1} \big(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} - \Omega \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} - \mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Omega + n^2 \Omega \otimes \Omega \big). \tag{7.233}$$

Proof. Let $V \in \mathrm{U}(\mathcal{X} \otimes \mathcal{X} \otimes \mathcal{X} \otimes \mathcal{X})$ be the permutation operator defined by the equation

$$V \operatorname{vec}(Y \otimes Z) = \operatorname{vec}(Y) \otimes \operatorname{vec}(Z), \tag{7.234}$$

holding for all $Y, Z \in \mathrm{L}(\mathcal{X})$. Alternatively, this operator may be defined by the equation

$$V(x_1 \otimes x_2 \otimes x_3 \otimes x_4) = x_1 \otimes x_3 \otimes x_2 \otimes x_4 \tag{7.235}$$

holding for all $x_1, x_2, x_3, x_4 \in \mathcal{X}$. As V is its own inverse, one has

$$V\big(\operatorname{vec}(Y) \otimes \operatorname{vec}(Z)\big) = \operatorname{vec}(Y \otimes Z) \tag{7.236}$$

for all $Y, Z \in \mathrm{L}(\mathcal{X})$. For every choice of maps $\Phi_0, \Phi_1 \in \mathrm{T}(\mathcal{X})$, it holds that

$$VJ(\Phi_0 \otimes \Phi_1)V^* = J(\Phi_0) \otimes J(\Phi_1). \tag{7.237}$$

Now, the Choi representation of Ξ is given by

$$J(\Xi) = \int \operatorname{vec}(U) \operatorname{vec}(U)^* \otimes \operatorname{vec}\big(\overline{U}\big) \operatorname{vec}\big(\overline{U}\big)^* \mathrm{d}\eta(U), \tag{7.238}$$

and therefore

$$VJ(\Xi)V^* = \int \operatorname{vec}\big(U \otimes \overline{U}\big) \operatorname{vec}\big(U \otimes \overline{U}\big)^* \mathrm{d}\eta(U). \tag{7.239}$$

This operator is the Choi representation of the isotropic twirling channel

$$\Lambda(X) = \int (U \otimes \overline{U}) X (U \otimes \overline{U})^* \, \mathrm{d}\eta(U) \tag{7.240}$$

defined in Example 7.26. From the analysis presented in that example, it follows that

$$\begin{aligned} VJ(\Xi)V^* &= \frac{1}{n^2} J(\mathbb{1}_{\mathrm{L}(\mathcal{X})}) \otimes J(\mathbb{1}_{\mathrm{L}(\mathcal{X})}) \\ &\quad + \frac{1}{n^2 - 1} \left(nJ(\Omega) - \frac{1}{n} J(\mathbb{1}_{\mathrm{L}(\mathcal{X})}) \right) \otimes \left(nJ(\Omega) - \frac{1}{n} J(\mathbb{1}_{\mathrm{L}(\mathcal{X})}) \right). \end{aligned} \tag{7.241}$$

By expanding the expression (7.241) and making use of the identity (7.237), one obtains (7.233), as required. □

Theorem 7.32. *Let $\mathcal{X}$ be a complex Euclidean space with dimension $n \geq 2$, and let $\Omega \in \mathrm{C}(\mathcal{X})$ denote the completely depolarizing channel on $\mathcal{X}$. For every unital channel $\Phi \in \mathrm{C}(\mathcal{X})$, it holds that*

$$\frac{n^2 - 2}{n^2 - 1} \Omega + \frac{1}{n^2 - 1} \Phi \tag{7.242}$$

is a mixed-unitary channel.

Proof. Let $\Psi \in \mathrm{CP}(\mathcal{X})$ be the map defined as

$$\Psi(X) = \int \big\langle \mathrm{vec}(U)\, \mathrm{vec}(U)^*, J(\Phi) \big\rangle \, UXU^* \, \mathrm{d}\eta(U), \tag{7.243}$$

for η being the Haar measure on $\mathrm{U}(\mathcal{X})$. It holds that

$$\int \mathrm{vec}(U)\, \mathrm{vec}(U)^* \, \mathrm{d}\eta(U) = \frac{1}{n} \mathbb{1}_{\mathcal{X} \otimes \mathcal{X}}, \tag{7.244}$$

and therefore

$$\int \big\langle \mathrm{vec}(U)\, \mathrm{vec}(U)^*, J(\Phi) \big\rangle \, \mathrm{d}\eta(U) = \frac{1}{n} \mathrm{Tr}(J(\Phi)) = 1. \tag{7.245}$$

It follows that the mapping Ψ is a mixed-unitary channel.

By Lemma 7.31, one has $J(\Psi) = \Xi(J(\Phi))$ for $\Xi \in \mathrm{CP}(\mathcal{X} \otimes \mathcal{X})$ being defined as

$$\Xi = \frac{1}{n^2 - 1} \big(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} - \Omega \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})} - \mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Omega + n^2 \Omega \otimes \Omega \big). \tag{7.246}$$

By the assumption that Φ is a unital channel, it holds that

$$\begin{aligned}(\Omega \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})(J(\Phi)) &= (\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Omega)(J(\Phi)) \\ &= (\Omega \otimes \Omega)(J(\Phi)) = \frac{\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{X}}}{n},\end{aligned} \tag{7.247}$$

and therefore

$$J(\Psi) = \frac{1}{n^2 - 1} J(\Phi) + \frac{n^2 - 2}{n(n^2 - 1)} \mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{X}}. \tag{7.248}$$

This is equivalent to Ψ being equal to (7.242), and therefore completes the proof. □

Corollary 7.33. *Let $\mathcal{X}$ be a complex Euclidean space having dimension $n \geq 2$, let $\Omega \in \mathrm{C}(\mathcal{X})$ denote the completely depolarizing channel on $\mathcal{X}$, and let $\Phi \in \mathrm{T}(\mathcal{X})$ be a trace-preserving and unital map satisfying*

$$\| J(\Omega) - J(\Phi) \| \leq \frac{1}{n(n^2 - 1)}. \tag{7.249}$$

It holds that Φ is a mixed-unitary channel.

Proof. Define a map $\Psi \in \mathrm{T}(\mathcal{X})$ as

$$\Psi = (n^2 - 1)\Phi - (n^2 - 2)\Omega. \tag{7.250}$$

It holds that Ψ is trace-preserving and unital. Moreover, one has

$$\begin{aligned}J(\Psi) &= (n^2 - 1)(J(\Phi) - J(\Omega)) + J(\Omega) \\ &= (n^2 - 1)(J(\Phi) - J(\Omega)) + \frac{1}{n} \mathbb{1}_{\mathcal{X} \otimes \mathcal{X}},\end{aligned} \tag{7.251}$$

which, by the assumptions of the corollary, implies that Ψ is completely positive. By Theorem 7.32 it follows that

$$\frac{n^2 - 2}{n^2 - 1} \Omega + \frac{1}{n^2 - 1} \Psi = \Phi \tag{7.252}$$

is a mixed-unitary channel, which completes the proof. □

7.3 Measure concentration and it applications

The unitarily invariant measures introduced in the previous section exhibit a phenomenon known as *measure concentration*.[4] In the case of the uniform spherical measure μ defined on the unit sphere of a complex Euclidean space $\mathcal{X}$, this phenomenon is reflected by the fact that, for every Lipschitz continuous function $f : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$, the subset of $\mathcal{S}(\mathcal{X})$ on which f differs significantly from its average value (or, alternately, its median values) must have relatively small measure. This phenomenon becomes more and more pronounced as the dimension of $\mathcal{X}$ grows.

Measure concentration is particularly useful in the theory of quantum information when used in the context of the *probabilistic method*. Various objects of interest, such as channels possessing certain properties, may be shown to exist by considering random choices of these object (typically based on the uniform spherical measure or Haar measure), followed by an analysis that demonstrates that the randomly chosen object possesses the property of interest with a nonzero probability. This method has been used successfully to demonstrate the existence of several interesting classes of objects for which explicit constructions are not known.

The present section explains this methodology, with its primary goal being to prove that the minimum output entropy of quantum channels is non-additive. Toward this goal, concentration bounds are established for uniform spherical measures, leading to an asymptotically strong form of a theorem known as *Dvoretzky's theorem*.

7.3.1 Lévy's lemma and Dvoretzky's theorem

The present subsection establishes facts concerning the concentration of measure phenomenon mentioned above for the measures defined in the previous section. A selection of bounds will be presented, mainly targeted toward a proof of Dvoretzky's theorem, which concerns the existence of a relatively large subspace $\mathcal{V}$ of a given complex Euclidean space $\mathcal{X}$ on which a given Lipschitz function $f : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$ *never* deviates significantly from its mean or median values with respect to the uniform spherical measure.

[4] Measure concentration is not limited to the measures introduced in the previous section—it is a more general phenomenon. For the purposes of this book, however, it will suffice to consider measure concentration with respect to those particular measures.

Concentration bounds for Gaussian measure

In order to prove concentration bounds for the uniform spherical measure, with respect to a given complex Euclidean space $\mathcal{X}$, it is helpful to begin by proving an analogous result for the standard Gaussian measure on $\mathbb{R}^n$. Theorem 7.34, which is stated and proved below, establishes a result of this form that serves as a starting point for the concentration bounds to follow.

In the statements of the theorems representing concentration bounds to be presented below, including Theorem 7.34, it will be necessary to refer to certain universal real number constants. Such constants will, as a general convention, be denoted δ, δ_1, δ_2, etc., and must be chosen to be sufficiently small for the various theorems to hold. Although the optimization of these absolute constants should not be seen as being necessarily uninteresting or unimportant, this goal will be considered as being secondary in this book. Suitable values for these constants will be given in each case, but in some cases these values have been selected to simplify expressions and proofs rather than to optimize their values.

Theorem 7.34. *There exists a positive real number $\delta_1 > 0$ for which the following holds. For every choice of a positive integer n, a κ-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, and a positive real number $\varepsilon > 0$, it holds that*

$$\Pr\bigl(f(X_1, \ldots, X_n) - \mathrm{E}(f(X_1, \ldots, X_n)) \geq \varepsilon\bigr) \leq \exp\left(-\frac{\delta_1 \varepsilon^2}{\kappa^2}\right), \tag{7.253}$$

for $X_1, \ldots, X_n$ being independent and identically distributed standard normal random variables.

Remark 7.35. One may take $\delta_1 = 2/\pi^2$.

The proof of Theorem 7.34 will make use of the two lemmas that follow. The first lemma is a fairly standard smoothing argument that will allow for basic multivariate calculus to be applied in the proof of the theorem.

Lemma 7.36. *Let n be a positive integer, let $f : \mathbb{R}^n \to \mathbb{R}$ be a κ-Lipschitz function, and let $\varepsilon > 0$ be a positive real number. There exists a differentiable κ-Lipschitz function $g : \mathbb{R}^n \to \mathbb{R}$ such that $|f(x) - g(x)| \leq \varepsilon$ for every $x \in \mathbb{R}^n$.*

Proof. For every $\delta > 0$, define a function $g_\delta : \mathbb{R}^n \to \mathbb{R}$ as

$$g_\delta(x) = \int f(x + \delta z) \, \mathrm{d}\gamma_n(z) \tag{7.254}$$

for all $x \in \mathbb{R}^n$, where γ_n denotes the standard Gaussian measure on $\mathbb{R}^n$. It will be proved that setting $g = g_\delta$ for a suitable choice of δ satisfies the requirements of the lemma.

First, by the assumption that f is κ-Lipschitz, it holds that

$$\begin{aligned} |f(x) - g_\delta(x)| &\leq \int |f(x) - f(x + \delta z)| \, \mathrm{d}\gamma_n(z) \\ &\leq \delta\kappa \int \|z\| \, \mathrm{d}\gamma_n(z) \leq \delta\kappa\sqrt{n} \end{aligned} \tag{7.255}$$

for all $x \in \mathbb{R}^n$ and $\delta > 0$. The last inequality in (7.255) makes use of (1.283) in Chapter 1. At this point, one may fix

$$\delta = \frac{\varepsilon}{\kappa\sqrt{n}} \tag{7.256}$$

and $g = g_\delta$, so that $|f(x) - g(x)| \leq \varepsilon$ for every $x \in \mathbb{R}^n$.

Next, it holds that g is κ-Lipschitz, as the following calculation shows:

$$\begin{aligned} |g(x) - g(y)| &\leq \int |f(x + \delta z) - f(y + \delta z)| \, \mathrm{d}\gamma_n(z) \\ &\leq \int \kappa \|x - y\| \, \mathrm{d}\gamma_n(z) = \kappa\|x - y\|, \end{aligned} \tag{7.257}$$

for every $x, y \in \mathbb{R}^n$.

It remains to prove that g is differentiable. Using the definition of the standard Gaussian measure, one may calculate that the gradient of g at an arbitrary point $x \in \mathbb{R}^n$ is given by

$$\nabla g(x) = \int z f(x + \delta z) \, \mathrm{d}\gamma_n(z). \tag{7.258}$$

The fact that the integral on the right-hand side of (7.258) exists follows from the inequality

$$\begin{aligned} \int \big\| z f(x + \delta z) \big\| \, \mathrm{d}\gamma_n(z) &\leq \int \kappa\big(\|z\|\|x\| + \delta\|z\|^2\big) \, \mathrm{d}\gamma_n(z) \\ &\leq \kappa\|x\|\sqrt{n} + \kappa\delta n. \end{aligned} \tag{7.259}$$

Moreover, it holds that $\nabla g(x)$ is a continuous function of x (and in fact is Lipschitz continuous), as

$$\begin{aligned} \big\| \nabla g(x) - \nabla g(y) \big\| &\leq \int \|z\| |f(x + \delta z) - f(y + \delta z)| \, \mathrm{d}\gamma_n(z) \\ &\leq \kappa\|x - y\|\sqrt{n}. \end{aligned} \tag{7.260}$$

As $\nabla g(x)$ is a continuous function of x, it follows that g is differentiable, which completes the proof. □

The second lemma establishes that the random variable $f(X_1, \ldots, X_n)$, for independent and normally distributed random variables $X_1, \ldots, X_n$ and a differentiable κ-Lipschitz function f, does not deviate too much from an independent copy of itself.

Lemma 7.37. *Let n be a positive integer, let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function satisfying $\|\nabla f(x)\| \leq \kappa$ for every $x \in \mathbb{R}^n$, and let*

$$X = (X_1, \ldots, X_n) \quad \text{and} \quad Y = (Y_1, \ldots, Y_n) \tag{7.261}$$

be vector-valued random variables, for $X_1, \ldots, X_n, Y_1, \ldots, Y_n$ being independent and identically distributed stamdard normal random variables. It holds that

$$\mathrm{E}\big(\exp(\lambda f(X) - \lambda f(Y))\big) \leq \exp\left(\frac{\lambda^2 \pi^2 \kappa^2}{8}\right) \tag{7.262}$$

for every real number $\lambda \in \mathbb{R}$.

Proof. First, define a function $g_{x,y} : \mathbb{R} \to \mathbb{R}$, for every choice of vectors $x, y \in \mathbb{R}^n$, as follows:

$$g_{x,y}(\theta) = f(\sin(\theta)x + \cos(\theta)y). \tag{7.263}$$

Applying the chain rule for differentiation, one finds that

$$g'_{x,y}(\theta) = \big\langle \nabla f(\sin(\theta)x + \cos(\theta)y), \cos(\theta)x - \sin(\theta)y \big\rangle \tag{7.264}$$

for every $x, y \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$. By the fundamental theorem of calculus, it therefore follows that

$$\begin{aligned} f(x) - f(y) &= g_{x,y}(\pi/2) - g_{x,y}(0) = \int_0^{\frac{\pi}{2}} g'_{x,y}(\theta) \mathrm{d}\theta \\ &= \int_0^{\frac{\pi}{2}} \big\langle \nabla f(\sin(\theta)x + \cos(\theta)y), \cos(\theta)x - \sin(\theta)y \big\rangle \, \mathrm{d}\theta. \end{aligned} \tag{7.265}$$

Next, define a random variable Z_θ, for each $\theta \in [0, \pi/2]$, as

$$Z_\theta = \big\langle \nabla f(\sin(\theta)X + \cos(\theta)Y), \cos(\theta)X - \sin(\theta)Y \big\rangle. \tag{7.266}$$

By (7.265), it follows that

$$\mathrm{E}\big(\exp(\lambda f(X) - \lambda f(Y))\big) = \mathrm{E}\left(\exp\left(\lambda \int_0^{\frac{\pi}{2}} Z_\theta \, \mathrm{d}\theta\right)\right). \tag{7.267}$$

By Jensen's inequality, one has

$$\mathrm{E}\left(\exp\left(\lambda \int_0^{\frac{\pi}{2}} Z_\theta \, \mathrm{d}\theta\right)\right) \leq \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \mathrm{E}\left(\exp\left(\frac{\pi\lambda}{2} Z_\theta\right)\right) \mathrm{d}\theta. \tag{7.268}$$

Finally, one arrives at a key step of the proof: the observation that each of the random variables Z_θ is identically distributed, as a consequence of the invariance of Gaussian measure under orthogonal transformations. That is, one has the following equality of vector-valued random variables:

$$\begin{pmatrix} \sin(\theta) X + \cos(\theta) Y \\ \cos(\theta) X - \sin(\theta) Y \end{pmatrix} = \begin{pmatrix} \sin(\theta)\mathbb{1} & \cos(\theta)\mathbb{1} \\ \cos(\theta)\mathbb{1} & -\sin(\theta)\mathbb{1} \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}. \tag{7.269}$$

As the distribution of $(X, Y) = (X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ is invariant under orthogonal transformations, it follows that the distribution of Z_θ does not depend on θ. Consequently,

$$\frac{2}{\pi} \int_0^{\frac{\pi}{2}} \mathrm{E}\left(\exp\left(\frac{\pi\lambda}{2} Z_\theta\right)\right) \mathrm{d}\theta = \mathrm{E}\left(\exp\left(\frac{\pi\lambda}{2} Z_0\right)\right). \tag{7.270}$$

This quantity can be evaluated using the Gaussian integral equation (1.272), yielding

$$\mathrm{E}\left(\exp\left(\frac{\pi\lambda}{2} Z_0\right)\right) = \mathrm{E}\left(\exp\left(\frac{\pi^2\lambda^2}{8} \| \nabla f(Y) \|^2\right)\right). \tag{7.271}$$

As it is to be assumed that $\| \nabla f(x) \| \leq \kappa$ for all $x \in \mathbb{R}^n$, the required bound is obtained as a result of (7.267), (7.268), (7.270), and (7.271). □

Proof of Theorem 7.34. Let X be a vector-valued random variable, defined as $X = (X_1, \ldots, X_n)$, and let $\lambda > 0$ be a positive real number to be specified shortly. By Markov's inequality, one has

$$\begin{aligned} &\Pr\big(f(X) - \mathrm{E}(f(X)) \geq \varepsilon\big) \\ &\quad = \Pr\big(\exp\big(\lambda f(X) - \lambda\, \mathrm{E}(f(X))\big) \geq \exp(\lambda\varepsilon)\big) \\ &\quad \leq \exp(-\lambda\varepsilon)\, \mathrm{E}\big(\exp\big(\lambda f(X) - \lambda\, \mathrm{E}(f(X))\big)\big). \end{aligned} \tag{7.272}$$

By introducing a new random variable $Y = (Y_1, \ldots, Y_n)$, which is to be independent and identically distributed to X, one finds that

$$\mathrm{E}\big(\exp\big(\lambda f(X) - \lambda\,\mathrm{E}(f(X))\big)\big) \leq \mathrm{E}\big(\exp\big(\lambda f(X) - \lambda f(Y)\big)\big) \tag{7.273}$$

by Jensen's inequality. Combining the two previous equations yields

$$\Pr\big(f(X) - \mathrm{E}(f(X)) \geq \varepsilon\big) \leq \exp(-\lambda\varepsilon)\,\mathrm{E}\big(\exp\big(\lambda f(X) - \lambda f(Y)\big)\big). \tag{7.274}$$

Assume first that f is differentiable, so that $\|\nabla f(x)\| \leq \kappa$ for all $x \in \mathbb{R}^n$ by the assumption that f is κ-Lipschitz. By Lemma 7.37, it follows that

$$\exp(-\lambda\varepsilon)\,\mathrm{E}\big(\exp\big(\lambda f(X) - \lambda f(Y)\big)\big) \leq \exp\left(-\lambda\varepsilon + \frac{\lambda^2\pi^2\kappa^2}{8}\right). \tag{7.275}$$

Setting $\lambda = 4\varepsilon/(\pi^2\kappa^2)$, and combining (7.274) with (7.275), yields

$$\Pr\big(f(X) - \mathrm{E}(f(X)) \geq \varepsilon\big) \leq \exp\left(-\frac{2\varepsilon^2}{\pi^2\kappa^2}\right), \tag{7.276}$$

which is the bound claimed in the statement of the theorem (for $\delta_1 = 2/\pi^2$).

Finally, suppose that f is κ-Lipschitz, but not necessarily differentiable. By Lemma 7.36, for every choice of $\zeta \in (0, \varepsilon)$ there exists a differentiable κ-Lipschitz function $g : \mathbb{R}^n \to \mathbb{R}$ satisfying $|f(x) - g(x)| \leq \zeta$ for every $x \in \mathbb{R}^n$, and therefore

$$\Pr\big(f(X) - \mathrm{E}(f(X)) \geq \varepsilon\big) \leq \Pr\big(g(X) - \mathrm{E}(g(X)) \geq \varepsilon - 2\zeta\big). \tag{7.277}$$

Applying the above analysis to g in place of f therefore yields

$$\Pr\big(f(X) - \mathrm{E}(f(X)) \geq \varepsilon\big) \leq \exp\left(-\frac{2(\varepsilon - 2\zeta)^2}{\pi^2\kappa^2}\right). \tag{7.278}$$

As this inequality holds for every $\zeta \in (0, \varepsilon)$, the theorem follows. □

The following example illustrates the application of Theorem 7.34 to the Euclidean norm. The analysis to be presented in this example is relevant to the discussion of the uniform spherical measure to be discussed shortly.

Example 7.38. Let n be a positive integer and define $f(x) = \|x\|$ for each $x \in \mathbb{R}^n$. It is an immediate consequence of the triangle inequality that f is 1-Lipschitz:

$$\big|f(x) - f(y)\big| = \big|\|x\| - \|y\|\big| \leq \|x - y\| \tag{7.279}$$

for all $x, y \in \mathbb{R}^n$. The mean value of $f(X_1, \ldots, X_n)$, for $X_1, \ldots, X_n$ being independent and identically distributed standard normal random variables, has the following closed-form expression (q.v. Section 1.2.3):

$$\mathrm{E}\big(f(X_1, \ldots, X_n)\big) = \frac{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}. \tag{7.280}$$

From this expression, an analysis reveals that

$$\mathrm{E}\big(f(X_1, \ldots, X_n)\big) = v_n \sqrt{n}, \tag{7.281}$$

where $v_1, v_2, v_3, \ldots$ is a strictly increasing sequence that begins

$$v_1 = \sqrt{\frac{2}{\pi}}, \quad v_2 = \frac{\sqrt{\pi}}{2}, \quad v_3 = \sqrt{\frac{8}{3\pi}}, \quad \ldots \tag{7.282}$$

and converges to 1 in the limit as n goes to infinity.

For any positive real number $\varepsilon > 0$, one may conclude the following two bounds from Theorem 7.34:

$$\begin{aligned} \Pr\big(\big\|(X_1, \ldots, X_n)\big\| \leq (v_n - \varepsilon)\sqrt{n}\big) &\leq \exp\big(-\delta_1 \varepsilon^2 n\big), \\ \Pr\big(\big\|(X_1, \ldots, X_n)\big\| \geq (v_n + \varepsilon)\sqrt{n}\big) &\leq \exp\big(-\delta_1 \varepsilon^2 n\big). \end{aligned} \tag{7.283}$$

Consequently, one has

$$\Pr\big(\big|\big\|(X_1, \ldots, X_n)\big\| - v_n\sqrt{n}\big| \geq \varepsilon\sqrt{n}\big) \leq 2\exp\big(-\delta_1 \varepsilon^2 n\big). \tag{7.284}$$

This bound illustrates that The Euclidean norm of a vector $x \in \mathbb{R}^n$ is tightly concentrated around its mean value $v_n\sqrt{n}$.

Concentration bounds for uniform spherical measure

Given that the uniform spherical measure is derived from the standard Gaussian measure, as described in Section 7.2.1, it is not unreasonable to expect that Theorem 7.34 might lead to an analogous fact holding for the uniform spherical measure. Indeed this is the case, as the theorems to be presented below establish.

The first theorem concerns the deviation of a Lipschitz random variable, defined with respect to the uniform spherical measure, from its mean value.

Theorem 7.39 (Lévy's lemma, mean value form)**.** *There exists a positive real number $\delta_2 > 0$ for which the following holds. For every complex Euclidean space $\mathcal{X}$, every κ-Lipschitz random variable $X : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$, distributed with respect to the uniform spherical measure μ on $\mathcal{S}(\mathcal{X})$, and every positive real number $\varepsilon > 0$, it holds that*

$$\begin{aligned} \Pr\big(X - \mathrm{E}(X) \geq \varepsilon\big) &\leq 2\exp\left(-\frac{\delta_2 \varepsilon^2 n}{\kappa^2}\right), \\ \Pr\big(X - \mathrm{E}(X) \leq -\varepsilon\big) &\leq 2\exp\left(-\frac{\delta_2 \varepsilon^2 n}{\kappa^2}\right), \end{aligned} \tag{7.285}$$

and

$$\Pr\big(|X - \mathrm{E}(X)| \geq \varepsilon\big) \leq 3\exp\left(-\frac{\delta_2 \varepsilon^2 n}{\kappa^2}\right), \tag{7.286}$$

for $n = \dim(\mathcal{X})$.

Remark 7.40. One may take $\delta_2 = 1/(25\pi)$.

The proof of Lemma 7.39 will make use of the following lemma, which provides a simple mechanism for extending a Lipschitz function defined on the unit sphere of $\mathbb{C}^n$ to a Lipschitz function defined on all of $\mathbb{R}^{2n}$.

Lemma 7.41. *Let n be a positive integer and let $f : \mathcal{S}(\mathbb{C}^n) \to \mathbb{R}$ be a κ-Lipschitz function that is neither strictly positive nor strictly negative. Define a function $g : \mathbb{R}^{2n} \to \mathbb{R}$ as*

$$g(x, y) = \begin{cases} \|x + iy\| f\left(\frac{x+iy}{\|x+iy\|}\right) & \text{if } x + iy \neq 0 \\ 0 & \text{if } x + iy = 0 \end{cases} \tag{7.287}$$

for all $x, y \in \mathbb{R}^n$. It holds that g is a (3κ)-Lipschitz function.

Proof. The assumption that f is neither strictly positive nor strictly negative implies, for every unit vector $u \in \mathbb{C}^n$, that there must exist a unit vector $v \in \mathbb{C}^n$ such that $f(u)f(v) \leq 0$, which in turn implies

$$|f(u)| \leq |f(u) - f(v)| \leq \kappa \|u - v\| \leq 2\kappa, \tag{7.288}$$

by the assumption that f is κ-Lipschitz.

Now suppose that $x_0, y_0, x_1, y_1 \in \mathbb{R}^n$ are vectors. If it is the case that $x_0 + iy_0 = 0$ and $x_1 + iy_1 = 0$, then it is immediate that

$$|g(x_0, y_0) - g(x_1, y_1)| = 0. \tag{7.289}$$

If it holds that $x_0 + iy_0 \neq 0$ and $x_1 + iy_1 = 0$, then (7.288) implies

$$|g(x_0, y_0) - g(x_1, y_1)| = |g(x_0, y_0)| \leq 2\kappa \|x + iy\| = 2\kappa \|(x_0, y_0)\|. \quad (7.290)$$

A similar bound holds for the case in which $x_0 + iy_0 = 0$ and $x_1 + iy_1 \neq 0$.

Finally, suppose that $x_0 + iy_0$ and $x_1 + iy_1$ are both nonzero. Write

$$z_0 = x_0 + iy_0 \quad \text{and} \quad z_1 = x_1 + iy_1, \quad (7.291)$$

and set

$$\alpha_0 = \frac{1}{\|z_0\|} \quad \text{and} \quad \alpha_1 = \frac{1}{\|z_1\|}. \quad (7.292)$$

This implies that $\alpha_0 z_0$ and $\alpha_1 z_1$ are unit vectors. There is no loss of generality in assuming $\alpha_0 \leq \alpha_1$; the case in which $\alpha_1 \leq \alpha_0$ is handled in a symmetric manner. By the triangle inequality, one has

$$\begin{aligned} |g(x_0, y_0) - g(x_1, y_1)| &= \big| \|z_0\| f(\alpha_0 z_0) - \|z_1\| f(\alpha_1 z_1) \big| \\ &\leq |f(\alpha_0 z_0)| \|z_0 - z_1\| + \|z_1\| |f(\alpha_0 z_0) - f(\alpha_1 z_1)|. \end{aligned} \quad (7.293)$$

Using (7.288), one finds that the first term in the final expression of (7.293) is bounded in the following straightforward manner:

$$|f(\alpha_0 z_0)| \|z_0 - z_1\| \leq 2\kappa \|z_0 - z_1\| = 2\kappa \|(x_0, y_0) - (x_1, y_1)\|. \quad (7.294)$$

To bound the second term, it may first be noted that

$$\|z_1\| |f(\alpha_0 z_0) - f(\alpha_1 z_1)| \leq \kappa \|z_1\| \|\alpha_0 z_0 - \alpha_1 z_1\|, \quad (7.295)$$

again by the assumption that f is κ-Lipschitz. Given that $0 < \alpha_0 \leq \alpha_1$, together with the fact that $\alpha_0 z_0$ and $\alpha_1 z_1$ are unit vectors, one finds that

$$\|\alpha_0 z_0 - \alpha_1 z_1\| \leq \|\alpha_1 z_0 - \alpha_1 z_1\| = \frac{\|z_0 - z_1\|}{\|z_1\|}, \quad (7.296)$$

and therefore

$$\kappa \|z_1\| \|\alpha_0 z_0 - \alpha_1 z_1\| \leq \kappa \|z_0 - z_1\| = \kappa \|(x_0, y_0) - (x_1, y_1)\|. \quad (7.297)$$

It follows that

$$|g(x_0, y_0) - g(x_1, y_1)| \leq 3\kappa \|(x_0, y_0) - (x_1, y_1)\|. \quad (7.298)$$

It has therefore been established that g is (3κ)-Lipschitz, as required. □

Proof of Theorem 7.39. The random variable $X - \mathrm{E}(X)$ has mean value 0, and is therefore neither strictly positive nor strictly negative. As X is κ-Lipschitz, so too is $X - \mathrm{E}(X)$, and so it follows that

$$\big|X - \mathrm{E}(X)\big| \leq 2\kappa, \tag{7.299}$$

as argued in the first paragraph of the proof of Lemma 7.41. The inequalities (7.285) and (7.286) therefore hold trivially when $\varepsilon > 2\kappa$. For this reason it will be assumed that $\varepsilon \leq 2\kappa$ for the remainder of the proof. It will also be assumed that $\mathcal{X} = \mathbb{C}^n$, for n being an arbitrary positive integer, which will simplify the notation used throughout the proof, and which causes no loss of generality.

Define a function $g : \mathbb{R}^{2n} \to \mathbb{R}$ as

$$g(y,z) = \begin{cases} \|y + iz\| \left(X\left(\frac{y+iz}{\|y+iz\|}\right) - \mathrm{E}(X) \right) & \text{if } y + iz \neq 0 \\ 0 & \text{if } y + iz = 0 \end{cases} \tag{7.300}$$

for all $y, z \in \mathbb{R}^n$, which is a (3κ)-Lipschitz function by Lemma 7.41. Let $Y = (Y_1, \ldots, Y_n)$ and $Z = (Z_1, \ldots, Z_n)$ be vector-valued random variables, for $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ being independent and identically distributed standard normal random variables, and define a random variable

$$W = g(Y, Z). \tag{7.301}$$

As $X - \mathrm{E}(X)$ has mean value 0, it is evident that $\mathrm{E}(W) = 0$ as well. Finally, by considering the definition of the uniform spherical measure, one finds that

$$\Pr(X - \mathrm{E}(X) \geq \varepsilon) = \Pr\big(W \geq \varepsilon \|Y + iZ\|\big). \tag{7.302}$$

The probability (7.302) may be upper-bounded through the use of the union bound:

$$\Pr\big(X - \mathrm{E}(X) \geq \varepsilon\big) \leq \Pr\Big(W \geq \varepsilon\lambda\sqrt{2n}\Big) + \Pr\Big(\|Y + iZ\| \leq \lambda\sqrt{2n}\Big) \tag{7.303}$$

for every choice of $\lambda > 0$. By Theorem 7.34 it holds that

$$\Pr\Big(W \geq \varepsilon\lambda\sqrt{2n}\Big) \leq \exp\left(-\frac{2\delta_1\varepsilon^2\lambda^2 n}{9\kappa^2}\right), \tag{7.304}$$

and, as established in Example 7.38, it holds that

$$\Pr\Big(\big\|Y+iZ\big\| \le \lambda\sqrt{2n}\Big) \le \exp\Big(-2\delta_1(\upsilon_{2n}-\lambda)^2 n\Big). \tag{7.305}$$

Setting

$$\lambda = \frac{3\kappa\upsilon_{2n}}{3\kappa+\varepsilon} \tag{7.306}$$

yields

$$\Pr\big(X \ge \mathrm{E}(X)+\varepsilon\big) \le 2\exp\left(-\frac{2\delta_1\varepsilon^2\upsilon_{2n}^2 n}{(3\kappa+\varepsilon)^2}\right) \le 2\exp\left(-\frac{\delta_1\pi\varepsilon^2 n}{50\kappa^2}\right), \tag{7.307}$$

where the second inequality makes use of the assumption $\varepsilon \le 2\kappa$, along with the observation that $\upsilon_{2n} \ge \upsilon_2 = \sqrt{\pi}/2$. As one may take $\delta_1 = 2/\pi^2$ in Theorem 7.34, the first inequality is therefore proved for $\delta_2 = 1/(25\pi)$.

The second and third inequalities are proved in essentially the same manner. In particular, one has

$$\begin{aligned}&\Pr\big(X-\mathrm{E}(X) \le -\varepsilon\big)\\ &\quad\le \Pr\Big(W \le -\varepsilon\lambda\sqrt{2n}\Big) + \Pr\Big(\big\|Y+iZ\big\| \le \lambda\sqrt{2n}\Big)\end{aligned} \tag{7.308}$$

and

$$\begin{aligned}&\Pr\big(|X-\mathrm{E}(X)| \ge \varepsilon\big)\\ &\qquad\le \Pr\Big(W \ge \varepsilon\lambda\sqrt{2n}\Big) + \Pr\Big(W \le -\varepsilon\lambda\sqrt{2n}\Big)\\ &\qquad\quad + \Pr\Big(\big\|Y+iZ\big\| \le \lambda\sqrt{2n}\Big),\end{aligned} \tag{7.309}$$

and again setting $\lambda = 3\kappa\upsilon_n/(3\kappa+\varepsilon)$ yields the required bounds. □

The second theorem on measure concentration for the uniform spherical measure, stated and proved below, is similar in spirit to Theorem 7.39, but it is concerned with the deviation of a Lipschitz random variable from its *median value*—or, more generally, from any of its *central values*—rather than its mean value. The next definition makes precise the notions of a median value and a central value of a random variable, after which the theorem is stated and proved.

Definition 7.42. Let X be a random variable and let β be a real number. It is said that β is a *median value* of X if and only if

$$\Pr(X \geq \beta) \geq \frac{1}{2} \quad \text{and} \quad \Pr(X \leq \beta) \geq \frac{1}{2}, \tag{7.310}$$

and it is said that β is a *central value* of X if and only if

$$\Pr(X \geq \beta) \geq \frac{1}{4} \quad \text{and} \quad \Pr(X \leq \beta) \geq \frac{1}{4}. \tag{7.311}$$

Theorem 7.43 (Lévy's lemma, central value form)**.** *There exists a positive real number $\delta_3 > 0$ for which the following holds. For every complex Euclidean space $\mathcal{X}$, every κ-Lipschitz random variable $X : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$, distributed with respect to the uniform spherical measure μ on $\mathcal{S}(\mathcal{X})$, every central value β of X, and every positive real number $\varepsilon > 0$, it holds that*

$$\Pr\big(|X - \mathrm{E}(X)| \geq \varepsilon\big) \leq 8 \exp\left(-\frac{\delta_3 \varepsilon^2 n}{\kappa^2}\right), \tag{7.312}$$

for $n = \dim(\mathcal{X})$.

Remark 7.44. One may take $\delta_3 = 1/(100\pi)$.

Proof. Let

$$\zeta = \sqrt{\frac{\ln(8)\kappa^2}{\delta_2 n}}, \tag{7.313}$$

for δ_2 being any positive real number for which Theorem 7.39 holds. By that theorem, one may conclude that the following two inequalities hold:

$$\Pr(X - \mathrm{E}(X) > \zeta) < 2 \exp\left(-\frac{\delta_2 \zeta^2 n}{\kappa^2}\right) = \frac{1}{4}, \tag{7.314}$$

$$\Pr(X - \mathrm{E}(X) < -\zeta) < 2 \exp\left(-\frac{\delta_2 \zeta^2 n}{\kappa^2}\right) = \frac{1}{4}. \tag{7.315}$$

From these inequalities, one concludes that $|\mathrm{E}(X) - \beta| \leq \zeta$.

Now, if it is the case that $\varepsilon \geq 2\zeta$, then Theorem 7.39 implies

$$\begin{aligned} \Pr\big(|X - \beta| \geq \varepsilon\big) &\leq \Pr\big(|X - \mathrm{E}(X)| \geq \varepsilon - \zeta\big) \\ &\leq \Pr\left(|X - \mathrm{E}(X)| \geq \frac{\varepsilon}{2}\right) \leq 3 \exp\left(-\frac{\delta_2 \varepsilon^2 n}{4\kappa^2}\right). \end{aligned} \tag{7.316}$$

On the other hand, if $\varepsilon < 2\zeta$, then one has

$$\exp\left(-\frac{\delta_2 \varepsilon^2 n}{4\kappa^2}\right) > \exp\left(-\frac{\delta_2 \zeta^2 n}{\kappa^2}\right) = \frac{1}{8}, \tag{7.317}$$

so it must trivially hold that

$$\Pr\big(|X - \beta| \geq \varepsilon\big) \leq 8 \exp\left(-\frac{\delta_2 \varepsilon^2 n}{4\kappa^2}\right). \tag{7.318}$$

The required bound (7.312) therefore holds in both cases, provided one takes $\delta_3 \leq \delta_2/4$. As Theorem 7.39 holds for $\delta_2 = 1/(25\pi)$, the bound (7.312) holds for $\delta_3 = 1/(100\pi)$. □

Dvoretzky's theorem

Dvoretzky's theorem, which plays a key role in the section following this one, establishes that a Lipschitz random variable, defined with respect to the uniform spherical measure for a given complex Euclidean space $\mathcal{X}$, must remain close to its central values *everywhere* on the unit sphere $\mathcal{S}(\mathcal{V})$, for some choice of a subspace $\mathcal{V} \subseteq \mathcal{X}$ having relatively large dimension. There are, in fact, multiple variants and generalizations of Dvoretzky's theorem; the variant to be considered in this book is specific to the unitarily invariant measures defined previously in the present chapter, and is applicable to *phase-invariant* functions, which are defined as follows.

Definition 7.45. Let $\mathcal{X}$ be a complex Euclidean space and let $f : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$ be a function. The function f is said to be a *phase-invariant* function if and only if it holds that $f(x) = f(e^{i\theta}x)$ for all $x \in \mathcal{S}(\mathcal{X})$ and $\theta \in \mathbb{R}$.

Theorem 7.46 (Dvoretzky's theorem)**.** *For every real number $\zeta \in (0,1)$, there exists a real number $\delta = \delta(\zeta) \in (0,1)$ for which the following holds. Let $\mathcal{X}$ be a complex Euclidean space, let*

$$X : \mathcal{S}(\mathcal{X}) \to \mathbb{R} \tag{7.319}$$

be a κ-Lipschitz, phase-invariant random variable, distributed with respect to the uniform spherical measure on $\mathcal{S}(\mathcal{X})$, let β be a central value of X, let $\varepsilon > 0$ be a positive real number, and let $\mathcal{V} \subseteq \mathcal{X}$ be a subspace with

$$1 \leq \dim(\mathcal{V}) \leq \frac{\delta \varepsilon^2}{\kappa^2} \dim(\mathcal{X}). \tag{7.320}$$

For each unit vector $v \in \mathcal{V}$, define a random variable

$$Y_v : \mathrm{U}(\mathcal{X}) \to \mathbb{R}, \tag{7.321}$$

distributed with respect to the Haar measure on $\mathrm{U}(\mathcal{X})$, as

$$Y_v(U) = X(Uv) \tag{7.322}$$

for every $U \in \mathrm{U}(\mathcal{X})$. It holds that

$$\Pr\big(|Y_v - \beta| \leq \varepsilon \text{ for every } v \in \mathcal{S}(\mathcal{V})\big) \geq 1 - \zeta. \tag{7.323}$$

The proof of Theorem 7.46 will make use of the two lemmas that follow.

Lemma 7.47. *Let $\mathcal{X}$ be a complex Euclidean space of dimension $n \geq 2$ and let $f : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$ be a κ-Lipschitz, phase-invariant function. For every unit vector $u \in \mathcal{S}(\mathcal{X})$, define a random variable $X_u : \mathrm{U}(\mathcal{X}) \to \mathbb{R}$, distributed with respect to the Haar measure η on $\mathrm{U}(\mathcal{X})$, as*

$$X_u(U) = f(Uu) \tag{7.324}$$

for all $U \in \mathrm{U}(\mathcal{X})$. For any pair of linearly independent unit vectors $u, v \in \mathcal{X}$ and every positive real number $\varepsilon > 0$, it holds that

$$\Pr\big(|X_u - X_v| \geq \varepsilon\big) \leq 3 \exp\left(-\frac{\delta_2 \varepsilon^2 (n-1)}{\kappa^2 \|u - v\|^2}\right), \tag{7.325}$$

for δ_2 being any positive real number satisfying the requirements of Theorem 7.39.

Proof. The lemma will first be proved in the special case in which $\langle u, v \rangle$ is a nonnegative real number. First, define

$$\lambda = \frac{1 + \langle u, v \rangle}{2}, \tag{7.326}$$

which satisfies $1/2 \leq \lambda < 1$ by the assumption that $\langle u, v \rangle$ is nonnegative and u and v are linearly independent. Set

$$x = \frac{u + v}{2\sqrt{\lambda}} \quad \text{and} \quad y = \frac{u - v}{2\sqrt{1 - \lambda}}, \tag{7.327}$$

so that x and y are orthonormal unit vectors for which

$$\begin{aligned} u &= \sqrt{\lambda}\, x + \sqrt{1 - \lambda}\, y, \\ v &= \sqrt{\lambda}\, x - \sqrt{1 - \lambda}\, y. \end{aligned} \tag{7.328}$$

Next, let $\mathcal{Y}$ be any complex Euclidean space having dimension $n-1$ and let $V \in \mathrm{U}(\mathcal{Y}, \mathcal{X})$ be any isometry for which $\mathrm{im}(V) \perp x$. For every $U \in \mathrm{U}(\mathcal{X})$, define a random variable $Y_U : \mathcal{S}(\mathcal{Y}) \to \mathbb{R}$, distributed with respect to the uniform spherical measure μ on $\mathcal{S}(\mathcal{Y})$, as

$$Y_U(y) = f\Big(U\Big(\sqrt{\lambda}x + \sqrt{1-\lambda}Vy\Big)\Big) - f\Big(U\Big(\sqrt{\lambda}x - \sqrt{1-\lambda}Vy\Big)\Big) \tag{7.329}$$

for every $y \in \mathcal{S}(\mathcal{Y})$. Using the triangle inequality, along with the fact that

$$\|u - v\| = 2\sqrt{1-\lambda}, \tag{7.330}$$

one may verify that each Y_U is $(\kappa\|u-v\|)$-Lipschitz and satisfies $\mathrm{E}(Y_U) = 0$. By Lévy's lemma (Theorem 7.39), it therefore holds that

$$\Pr\big(|Y_U| \geq \varepsilon\big) \leq 3\exp\left(-\frac{\delta_2 \varepsilon^2 (n-1)}{\kappa^2 \|u-v\|^2}\right), \tag{7.331}$$

for every $U \in \mathrm{U}(\mathcal{X})$.

Finally, define a random variable $Z : \mathrm{U}(\mathcal{X}) \times \mathcal{S}(\mathcal{Y}) \to \mathbb{R}$, distributed with respect to the product measure $\eta \times \mu$, as

$$Z(U, y) = Y_U(y) \tag{7.332}$$

for all $U \in \mathrm{U}(\mathcal{X})$ and $y \in \mathcal{S}(\mathcal{Y})$. Because the uniform spherical measure and Haar measure are both unitarily invariant, it follows that Z and $X_u - X_v$ are identically distributed. It therefore holds that

$$\begin{aligned} \Pr\big(|X_u - X_v| \geq \varepsilon\big) &= \Pr\big(|Z| \geq \varepsilon\big) \\ &= \int \Pr\big(|Y_U| \geq \varepsilon\big)\, \mathrm{d}\eta(U) \leq 3\exp\left(-\frac{\delta_2 \varepsilon^2 (n-1)}{\kappa^2 \|u-v\|^2}\right), \end{aligned} \tag{7.333}$$

which proves the lemma in the case that $\langle u, v\rangle$ is a nonnegative real number.

In the situation in which $\langle u, v\rangle$ is not a nonnegative real number, one may choose $\alpha \in \mathbb{C}$ with $|\alpha| = 1$ so that $\langle u, \alpha v\rangle$ is a nonnegative real number. By the assumption that f is phase invariant, it holds that $X_v = X_{\alpha v}$, and therefore

$$\begin{aligned} \Pr\big(|X_u - X_v| \geq \varepsilon\big) &= \Pr\big(|X_u - X_{\alpha v}| \geq \varepsilon\big) \\ &\leq 3\exp\left(-\frac{\delta_2 \varepsilon^2 (n-1)}{\kappa^2 \|u-\alpha v\|^2}\right), \end{aligned} \tag{7.334}$$

by the analysis above. As it necessarily holds that $\|u - \alpha v\| \leq \|u - v\|$, it follows that

$$\Pr\bigl(|X_u - X_v| \geq \varepsilon\bigr) \leq 3 \exp\left(-\frac{\delta_2 \varepsilon^2 (n-1)}{\kappa^2 \|u - v\|^2}\right) \tag{7.335}$$

for every $\varepsilon > 0$, which completes the proof. □

The next lemma bounds the mean value of the maximum of a collection of nonnegative random variables satisfying a property reminiscent of the bounds obtained for the concentration results presented above.

Lemma 7.48. *Let $N \geq 2$ be a positive integer, let K and θ be positive real numbers, and let $Y_1, \ldots, Y_N$ be nonnegative random variables for which*

$$\Pr(Y_k \geq \lambda) \leq K \exp\bigl(-\theta \lambda^2\bigr) \tag{7.336}$$

for every $k \in \{1, \ldots, N\}$ and every $\lambda \geq 0$. It holds that

$$\mathrm{E}\bigl(\max\{Y_1, \ldots, Y_N\}\bigr) \leq \sqrt{\frac{\ln(N)}{\theta}} + \frac{K}{\sqrt{2\theta}}. \tag{7.337}$$

Proof. As the random variables $Y_1, \ldots, Y_N$ take only nonnegative values, one may write

$$\mathrm{E}\bigl(\max\{Y_1, \ldots, Y_N\}\bigr) = \int_0^\infty \Pr\bigl(\max\{Y_1, \ldots, Y_N\} \geq \lambda\bigr)\, \mathrm{d}\lambda. \tag{7.338}$$

Splitting the integral into two parts, and making use of the fact that the probability of any event is at most 1, yields

$$\begin{aligned} &\mathrm{E}\bigl(\max\{Y_1, \ldots, Y_N\}\bigr) \\ &\quad \leq \sqrt{\frac{\ln(N)}{\theta}} + \int_{\sqrt{\frac{\ln(N)}{\theta}}}^\infty \Pr\bigl(\max\{Y_1, \ldots, Y_N\} \geq \lambda\bigr)\, \mathrm{d}\lambda. \end{aligned} \tag{7.339}$$

By the union bound, together with the assumption (7.336) on $Y_1, \ldots, Y_N$, one has

$$\int_{\sqrt{\frac{\ln(N)}{\theta}}}^\infty \Pr\bigl(\max\{Y_1, \ldots, Y_N\} \geq \lambda\bigr)\, \mathrm{d}\lambda \leq KN \int_{\sqrt{\frac{\ln(N)}{\theta}}}^\infty \exp(-\theta\lambda^2)\, \mathrm{d}\lambda. \tag{7.340}$$

As $\ln(2) > 1/2$, it holds that $\lambda\sqrt{2\theta} > 1$ for every choice of λ satisfying

$$\lambda \geq \sqrt{\frac{\ln(N)}{\theta}}, \tag{7.341}$$

and therefore

$$\int_{\sqrt{\frac{\ln(N)}{\theta}}}^{\infty} \exp(-\theta\lambda^2)\,\mathrm{d}\lambda \;\leq\; \int_{\sqrt{\frac{\ln(N)}{\theta}}}^{\infty} \lambda\sqrt{2\theta}\exp(-\theta\lambda^2)\,\mathrm{d}\lambda \;=\; \frac{1}{N\sqrt{2\theta}}. \tag{7.342}$$

The required inequality now follows from (7.339), (7.340), and (7.342). □

Proof of Theorem 7.46. It will be proved that any choice of $\delta > 0$ satisfying

$$\delta < \zeta^2\left(\frac{8}{\sqrt{\delta_3}} + \frac{90}{\sqrt{\delta_2}}\right)^{-2}, \tag{7.343}$$

for δ_2 and δ_3 being positive real numbers that satisfy the requirements of Theorem 7.39 and Theorem 7.43, respectively, fulfills the requirements of the theorem. The theorem is trivial in the case $n = 1$, as the phase invariance of X implies that X is constant in this case, and for this reason it will be assumed that $n \geq 2$ for the remainder of the proof.

Assume hereafter that $X : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$, β, ε, and $\mathcal{V}$, as in the statement of the theorem, have been given. By Markov's inequality, one has

$$\begin{aligned} &\Pr\bigl(\sup\bigl\{|Y_v - \beta| \,:\, v \in \mathcal{S}(\mathcal{V})\bigr\} \leq \varepsilon\bigr) \\ &\qquad \geq 1 - \frac{\mathrm{E}\bigl(\sup\bigl\{|Y_v - \beta| \,:\, v \in \mathcal{S}(\mathcal{V})\bigr\}\bigr)}{\varepsilon}, \end{aligned} \tag{7.344}$$

so the theorem will follow from a demonstration that

$$\mathrm{E}\bigl(\sup\bigl\{|Y_v - \beta| \,:\, v \in \mathcal{S}(\mathcal{V})\bigr\}\bigr) \leq \zeta\varepsilon. \tag{7.345}$$

Let $m = \dim(\mathcal{V})$, and for each nonnegative integer $k \in \mathbb{N}$, let $\mathcal{N}_k$ be a minimal (2^{-k+1})-net for $\mathcal{S}(\mathcal{V})$. It is evident that $|\mathcal{N}_0| = 1$, and for every $k \in \mathbb{N}$ it holds that

$$|\mathcal{N}_k| \leq \bigl(1 + 2^k\bigr)^{2m} \leq 4^{(k+1)m} \tag{7.346}$$

by Theorem 1.8. For each $v \in \mathcal{S}(\mathcal{V})$ and $k \in \mathbb{N}$, fix $z_k(v) \in \mathcal{N}_k$ to be any element of the set $\mathcal{N}_k$ for which the distance to v is minimized, which implies that

$$\|v - z_k(v)\| \leq 2^{-k+1}. \tag{7.347}$$

One may observe that $z_0 = z_0(v)$ is independent of v, as there is a single element in the set $\mathcal{N}_0$, and also that

$$\lim_{k\to\infty} z_k(v) = v \tag{7.348}$$

for every $v \in \mathcal{S}(\mathcal{V})$.

Next, observe that

$$X(Uv) = X(Uz_0) + \sum_{k=0}^{\infty} \Big(X\big(Uz_{k+1}(v)\big) - X\big(Uz_k(v)\big) \Big), \tag{7.349}$$

for every $v \in \mathcal{S}(\mathcal{V})$ and $U \in \mathrm{U}(\mathcal{X})$; this fact may be verified by telescoping the sum and making use of (7.348), along with the continuity of X. It follows that

$$Y_v = Y_{z_0} + \sum_{k=0}^{\infty} \big(Y_{z_{k+1}(v)} - Y_{z_k(v)} \big) \tag{7.350}$$

for every $v \in \mathcal{S}(\mathcal{V})$. By the triangle inequality, one therefore has

$$\begin{aligned} &\sup\big\{ |Y_v - \beta| \,:\, v \in \mathcal{S}(\mathcal{V}) \big\} \\ &\quad \leq |Y_{z_0} - \beta| + \sup\left\{ \sum_{k=0}^{\infty} \big| Y_{z_{k+1}(v)} - Y_{z_k(v)} \big| \,:\, v \in \mathcal{S}(\mathcal{V}) \right\}. \end{aligned} \tag{7.351}$$

The expeted value of the two terms on the right-hand side of this inequality will be bounded separately.

The expected value of the first term $|Y_{z_0} - \beta|$ will be considered first. By Theorem 7.23, the random variable Y_{z_0} is identically distributed to X. It follows by Theorem 7.43 that

$$\Pr\big(|Y_{z_0} - \beta| \geq \lambda\big) = \Pr\big(|X - \beta| \geq \lambda\big) \leq 8 \exp\left(-\frac{\delta_3 \lambda^2 n}{\kappa^2} \right) \tag{7.352}$$

for every $\lambda \geq 0$, which implies

$$\begin{aligned} \mathrm{E}\big(|Y_{z_0} - \beta|\big) &= \int_0^{\infty} \Pr\big(|Y_{z_0} - \beta| \geq \lambda\big)\, \mathrm{d}\lambda \\ &\leq 8 \int_0^{\infty} \exp\left(-\frac{\delta_3 \lambda^2 n}{\kappa^2} \right) \mathrm{d}\lambda = 4\sqrt{\frac{\pi \kappa^2}{\delta_3 n}} < \frac{8\kappa}{\sqrt{\delta_3 n}}. \end{aligned} \tag{7.353}$$

It remains to bound the expected value of the second term on the right-hand side of (7.351). It holds that

$$\|z_{k+1}(v) - z_k(v)\| \leq \|z_{k+1}(v) - v\| + \|v - z_k(v)\| < 2^{-k+2} \tag{7.354}$$

for all $v \in \mathcal{S}(\mathcal{V})$ and all $k \in \mathbb{N}$, and therefore

$$\begin{aligned} &\sup\left\{\sum_{k=0}^{\infty} \left|Y_{z_{k+1}(v)} - Y_{z_k(v)}\right| \,:\, v \in \mathcal{S}(\mathcal{V})\right\} \\ &\qquad \leq \sum_{k=0}^{\infty} \max\left\{|Y_x - Y_y| \,:\, (x,y) \in \mathcal{M}_k\right\} \end{aligned} \tag{7.355}$$

where

$$\mathcal{M}_k = \left\{(x,y) \in \mathcal{N}_{k+1} \times \mathcal{N}_k,\ \|x - y\| < 2^{-k+2}\right\}. \tag{7.356}$$

By Lemma 7.47, it holds that

$$\Pr\Big(\big|Y_x - Y_y\big| \geq \varepsilon\Big) \leq 3\exp\left(-\frac{\delta_2 \varepsilon^2 (n-1)}{\kappa^2 \|x - y\|^2}\right) \tag{7.357}$$

for every pair of linearly independent vectors $x, y \in \mathcal{S}(\mathcal{V})$, for δ_2 being any positive real number for which Theorem 7.39 holds. (By the assumption that X is phase-invariant, one has $Y_x = Y_y$ if $x, y \in \mathcal{S}(\mathcal{V})$ are linearly dependent.) For each choice of $k \in \mathbb{N}$, it therefore follows from Lemma 7.48 that

$$\mathrm{E}\Big(\max\Big\{|Y_x - Y_y| \,:\, (x,y) \in \mathcal{M}_k\Big\}\Big) \leq \sqrt{\frac{\ln(N)}{\theta}} + \frac{3}{\sqrt{2\theta}} \tag{7.358}$$

for

$$\theta = \frac{4^k \delta_2 (n-1)}{16\kappa^2} \quad \text{and} \quad N = |\mathcal{M}_k| < 16^{(k+2)m}. \tag{7.359}$$

The remainder of the proof consists of routine calculations showing that the required bound is achieved. Using the bound

$$\sqrt{\ln(N)} \leq \sqrt{\log(N)} < 2\sqrt{(k+2)m}, \tag{7.360}$$

summing over all $k \in \mathbb{N}$, and making use of the summations

$$\sum_{k=0}^{\infty} 2^{-k}\sqrt{k+2} < \frac{7}{2} \quad \text{and} \quad \sum_{k=0}^{\infty} 2^{-k} = 2, \tag{7.361}$$

one concludes that

$$\sum_{k=0}^{\infty} \mathrm{E}\Big(\max\Big\{|Y_x - Y_y| : (x,y) \in \mathcal{M}_k\Big\}\Big) < \frac{90\kappa}{\sqrt{\delta_2}}\sqrt{\frac{m}{n}}. \tag{7.362}$$

By (7.351), (7.353), and (7.362), it follows that

$$\mathrm{E}\bigl(\sup\bigl\{|Y_v - \beta| \,:\, v \in \mathcal{S}(\mathcal{V})\bigr\}\bigr) < \left(\frac{8}{\sqrt{\delta_3}} + \frac{90}{\sqrt{\delta_2}}\right)\kappa\sqrt{\frac{m}{n}}. \tag{7.363}$$

Under the assumption that

$$m \leq \frac{\delta\varepsilon^2 n}{\kappa^2}, \tag{7.364}$$

for δ satisfying (7.343), it therefore holds that

$$\mathrm{E}\bigl(\sup\bigl\{|Y_v - \beta| \,:\, v \in \mathcal{S}(\mathcal{V})\bigr\}\bigr) < \zeta\varepsilon, \tag{7.365}$$

which completes the proof. □

7.3.2 Applications of measure concentration

Two applications of the results on measure concentration discussed in the previous subsection will now be presented. The first is a demonstration that most pure states of a pair of registers are highly entangled, and the second is a proof that the minimum output entropy of channels is non-additive in general. The two applications are related, with the second depending on the first.

Most pure states are highly entangled

Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are complex Euclidean spaces, and suppose further that the dimensions $n = \dim(\mathcal{X})$ and $m = \dim(\mathcal{Y})$ of these spaces satisfy $n \leq m$. For some choices of a unit vector $u \in \mathcal{X} \otimes \mathcal{Y}$, it holds that

$$\mathrm{Tr}_{\mathcal{Y}}(uu^*) = \omega, \tag{7.366}$$

for $\omega = \mathbb{1}/n$ denoting the completely mixed state with respect to $\mathcal{X}$. Of course, not every unit vector $u \in \mathcal{X} \otimes \mathcal{Y}$ satisfies this equation (unless $n = 1$); but as n grows, the equation holds approximately for an increasingly large portion of the set $\mathcal{S}(\mathcal{X} \otimes \mathcal{Y})$.

The following lemma establishes one specific fact along these lines, in which an approximation with respect to the 2-norm distance between states is considered. The proof makes use of Lévy's lemma (Theorem 7.39), along with calculations of integrals involving the uniform spherical measure.

Lemma 7.49. *There exists a positive real number K_0 with the following property. For any complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, having dimensions $n = \dim(\mathcal{X})$ and $m = \dim(\mathcal{Y})$, and for $X : \mathcal{S}(\mathcal{X} \otimes \mathcal{Y}) \to \mathbb{R}$ being a random variable, distributed with respect to the uniform spherical measure on $\mathcal{S}(\mathcal{X} \otimes \mathcal{Y})$, defined as*

$$X(u) = \left\| \operatorname{Tr}_{\mathcal{Y}}(uu^*) - \omega \right\|_2 \tag{7.367}$$

for $\omega = \mathbb{1}/n$, it holds that

$$\Pr\left(X \geq \frac{K_0}{\sqrt{m}} \right) < 4^{-n}. \tag{7.368}$$

Proof. It will be proved that the lemma holds for $K_0 = \sqrt{12/\delta_2} + 1$, for δ_2 being any positive real number satisfying the requirements of the mean value form of Lévy's lemma (Theorem 7.39).

The random variable X may alternatively be defined as

$$X(\operatorname{vec}(A)) = \left\| AA^* - \omega \right\|_2 \tag{7.369}$$

for every operator $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ satisfying $\|A\|_2 = 1$. The triangle inequality implies that

$$\left| X(\operatorname{vec}(A)) - X(\operatorname{vec}(B)) \right| \leq \left\| AA^* - BB^* \right\|_2. \tag{7.370}$$

Again using the triangle inequality, along with the fact that the 2-norm is submultiplicative, one has

$$\begin{aligned} \left\| AA^* - BB^* \right\|_2 &\leq \left\| AA^* - AB^* \right\|_2 + \left\| AB^* - BB^* \right\|_2 \\ &\leq \big(\|A\|_2 + \|B\|_2 \big) \|A - B\|_2 \leq 2 \|A - B\|_2. \end{aligned} \tag{7.371}$$

It therefore holds that X is 2-Lipschitz.

Next, it will be proved that

$$\mathrm{E}(X) \leq \frac{1}{\sqrt{m}}. \tag{7.372}$$

This bound follows from Jensen's inequality,

$$\big(\mathrm{E}(X)\big)^2 \leq \mathrm{E}\big(X^2\big), \tag{7.373}$$

along with an evaluation of $\mathrm{E}(X^2)$. To evaluate this expectation, observe first that

$$\left\| \operatorname{Tr}_{\mathcal{Y}}(uu^*) - \omega \right\|_2^2 = \operatorname{Tr}\Big(\big(\operatorname{Tr}_{\mathcal{Y}}(uu^*)\big)^2 \Big) - \frac{1}{n}. \tag{7.374}$$

For every vector $u \in \mathcal{X} \otimes \mathcal{Y}$, it holds that

$$\operatorname{Tr}\Big(\big(\operatorname{Tr}_{\mathcal{Y}}(uu^*)\big)^2\Big) = \big\langle V, uu^* \otimes uu^* \big\rangle, \tag{7.375}$$

for $V \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y} \otimes \mathcal{X} \otimes \mathcal{Y})$ being the operator defined as

$$V(x_0 \otimes y_0 \otimes x_1 \otimes y_1) = x_1 \otimes y_0 \otimes x_0 \otimes y_1 \tag{7.376}$$

for all vectors $x_0, x_1 \in \mathcal{X}$ and $y_0, y_1 \in \mathcal{Y}$. Equivalently, for Σ and Γ denoting the alphabets for which $\mathcal{X} = \mathbb{C}^{\Sigma}$ and $\mathcal{Y} = \mathbb{C}^{\Gamma}$, one may write

$$V = \sum_{\substack{a,b \in \Sigma \\ c,d \in \Gamma}} E_{a,b} \otimes E_{c,c} \otimes E_{b,a} \otimes E_{d,d}. \tag{7.377}$$

Integrating with respect to the uniform spherical measure yields

$$\begin{aligned} \mathrm{E}\big(X^2\big) &= \int \big\langle V, uu^* \otimes uu^* \big\rangle \, \mathrm{d}\mu(u) - \frac{1}{n} \\ &= \frac{1}{\binom{nm+1}{2}} \big\langle V, \Pi_{(\mathcal{X}\otimes\mathcal{Y})\otimes(\mathcal{X}\otimes\mathcal{Y})} \big\rangle - \frac{1}{n}. \end{aligned} \tag{7.378}$$

A case analysis reveals that

$$\begin{aligned} &\big\langle E_{a,b} \otimes E_{c,c} \otimes E_{b,a} \otimes E_{d,d}, \Pi_{(\mathcal{X}\otimes\mathcal{Y})\otimes(\mathcal{X}\otimes\mathcal{Y})} \big\rangle \\ &= \begin{cases} 1 & \text{if } a = b \text{ and } c = d \\ \frac{1}{2} & \text{if } (a = b \text{ and } c \neq d) \text{ or } (a \neq b \text{ and } c = d) \\ 0 & \text{if } a \neq b \text{ and } c \neq d. \end{cases} \end{aligned} \tag{7.379}$$

Performing the required arithmetic yields

$$\mathrm{E}\big(X^2\big) = \frac{n+m}{nm+1} - \frac{1}{n} < \frac{1}{m}, \tag{7.380}$$

and therefore (7.372) has been established.

Finally, by the mean value form of Lévy's lemma (Theorem 7.39), one has

$$\Pr\left(X \geq \frac{K_0}{\sqrt{m}}\right) \leq 2 \exp\left(-\frac{\delta_2 (K_0 - 1)^2 n}{4}\right). \tag{7.381}$$

For $K_0 = \sqrt{12/\delta_2} + 1$, one has

$$2 \exp\left(-\frac{\delta_2 (K_0 - 1)^2 n}{4}\right) = 2 \exp(-3n) < 4^{-n}, \tag{7.382}$$

which completes the proof. □

If it is the case that $\mathrm{Tr}_{\mathcal{Y}}(uu^*)$ is approximately equal to the completely mixed state ω, for a given choice of a unit vector $u \in \mathcal{X} \otimes \mathcal{Y}$, then one might reasonably expect the entanglement entropy $\mathrm{H}\big(\mathrm{Tr}_{\mathcal{Y}}(uu^*)\big)$ of the pure state u to be approximately equal to its maximum possible value $\log(n)$, depending on the particular notions of approximate equality under consideration. The next lemma establishes a lower bound on the von Neumann entropy that allows a precise implication along these lines to be made when combined with Lemma 7.49.

Lemma 7.50. *Let $\mathcal{X}$ be a complex Euclidean space and let $n = \dim(\mathcal{X})$. For every density operator $\rho \in \mathrm{D}(\mathcal{X})$ it holds that*

$$\mathrm{H}(\rho) \geq \log(n) - \frac{n}{\ln(2)} \big\| \rho - \omega \big\|_2^2, \tag{7.383}$$

where $\omega = \mathbb{1}/n$ denotes the completely mixed state with respect to $\mathcal{X}$.

Proof. It holds that $\ln(\alpha) \leq \alpha - 1$ for all $\alpha > 0$, and therefore

$$\begin{aligned} \frac{n}{\ln(2)} \big\| \rho - \omega \big\|_2^2 &= \frac{n \,\mathrm{Tr}(\rho^2) - 1}{\ln(2)} \\ &\geq \log\big(n \,\mathrm{Tr}(\rho^2)\big) = \log(n) + \log\big(\mathrm{Tr}(\rho^2)\big). \end{aligned} \tag{7.384}$$

The logarithm function is concave, and therefore one has

$$-\mathrm{H}(p) = \sum_{a \in \Sigma} p(a) \log(p(a)) \leq \log\left(\sum_{a \in \Sigma} p(a)^2 \right) \tag{7.385}$$

for every alphabet Σ and every probability vector $p \in \mathcal{P}(\Sigma)$. Consequently,

$$-\mathrm{H}(\rho) \leq \log\big(\mathrm{Tr}\big(\rho^2\big)\big), \tag{7.386}$$

and therefore

$$\frac{n}{\ln(2)} \big\| \rho - \omega \big\|_2^2 \geq \log(n) - \mathrm{H}(\rho), \tag{7.387}$$

which is equivalent to the required inequality. □

As a consequence of Lemmas 7.49 and 7.50, it follows that most bipartite pure states have an entanglement entropy that is close to this quantity's maximum possible value.

Theorem 7.51. *There exists a positive real number K for which the following holds. For every choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, and for $X : \mathcal{S}(\mathcal{X} \otimes \mathcal{Y}) \to \mathbb{R}$ being a random variable, distributed with respect to the uniform spherical measure on $\mathcal{S}(\mathcal{X} \otimes \mathcal{Y})$ and defined as*

$$X(u) = \mathrm{H}\big(\mathrm{Tr}_{\mathcal{Y}}(uu^*)\big) \tag{7.388}$$

for every $u \in \mathcal{S}(\mathcal{X} \otimes \mathcal{Y})$, it holds that

$$\Pr\left(X \leq \log(n) - \frac{Kn}{m}\right) < 4^{-n} \tag{7.389}$$

for $n = \dim(\mathcal{X})$ and $m = \dim(\mathcal{Y})$.

Proof. It will be proved that the theorem holds for $K = K_0^2/\ln(2)$, where K_0 is any positive real number that satisfies the requirements of Lemma 7.49.

Define a random variable $Y : \mathcal{S}(\mathcal{X} \otimes \mathcal{Y}) \to \mathbb{R}$, distributed with respect to the uniform spherical measure, as

$$Y(u) = \big\| \mathrm{Tr}_{\mathcal{Y}}(uu^*) - \omega \big\|_2 \tag{7.390}$$

for every $u \in \mathcal{S}(\mathcal{X} \otimes \mathcal{Y})$. If a given unit vector $u \in \mathcal{X} \otimes \mathcal{Y}$ satisfies

$$Y(u) < \frac{K_0}{\sqrt{m}}, \tag{7.391}$$

then

$$X(u) > \log(n) - \frac{n}{\ln(2)} \frac{K_0^2}{m} = \log(n) - \frac{Kn}{m} \tag{7.392}$$

by Lemma 7.50. One therefore has that

$$\Pr\left(X > \log(n) - \frac{Kn}{m}\right) \geq \Pr\left(Y < \frac{K_0}{\sqrt{m}}\right) > 1 - 4^{-n} \tag{7.393}$$

by Lemma 7.49. This bound is equivalent to (7.389), which completes the proof. □

Counter-example to the additivity of minimum output entropy

The minimum output entropy of a channel is, as the following definition states explicitly, the minimum value of the von Neumann entropy that can be obtained by evaluating that channel on a quantum state input.

Definition 7.52. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. The *minimum output entropy* of Φ is defined as

$$\mathrm{H}_{\min}(\Phi) = \min\bigl\{\mathrm{H}(\Phi(\rho)) \,:\, \rho \in \mathrm{D}(\mathcal{X})\bigr\}. \tag{7.394}$$

It follows from the concavity of the von Neumann entropy function that the minimum output entropy $\mathrm{H}_{\min}(\Phi)$ of a given channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is achieved by a pure state:

$$\mathrm{H}_{\min}(\Phi) = \min\bigl\{\mathrm{H}(\Phi(uu^*)) \,:\, u \in \mathcal{S}(\mathcal{X})\bigr\}. \tag{7.395}$$

It was a long-standing conjecture that the minimum output entropy is additive with respect to tensor products of channels. The following theorem demonstrates that this is, in fact, not the case.

Theorem 7.53 (Hastings)**.** *There exist complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$ and channels $\Phi, \Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ such that*

$$\mathrm{H}_{\min}(\Phi \otimes \Psi) < \mathrm{H}_{\min}(\Phi) + \mathrm{H}_{\min}(\Psi). \tag{7.396}$$

A high-level overview of the proof of Theorem 7.53 is as follows. For each choice of a positive integer n, one may consider complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ with

$$\dim(\mathcal{X}) = n^2, \quad \dim(\mathcal{Y}) = n, \quad \text{and} \quad \dim(\mathcal{Z}) = n^2. \tag{7.397}$$

It will be proved, for a sufficiently large choice of n, that there exists an isometry $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ for which the channels $\Phi, \Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ defined as

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\bigl(VXV^*\bigr) \quad \text{and} \quad \Psi(X) = \mathrm{Tr}_{\mathcal{Z}}\bigl(\overline{V}XV^{\mathsf{T}}\bigr) \tag{7.398}$$

for all $X \in \mathrm{L}(\mathcal{X})$ yield the strict inequality (7.396). The existence of a suitable isometry V is proved using the probabilistic method: for any fixed isometry $V_0 \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$, the set of all unitary operators $U \in \mathrm{U}(\mathcal{Y} \otimes \mathcal{Z})$ for which the isometry $V = UV_0$ possesses the required property will be shown to have positive measure, with respect to the Haar measure on $\mathrm{U}(\mathcal{Y} \otimes \mathcal{Z})$.

The proof of Theorem 7.53 will make use of the lemmas that follow. The first lemma provides an upper bound on the minimum output entropy of the tensor product $\Phi \otimes \Psi$ for two channels Φ and Ψ defined as in (7.398).

Lemma 7.54. *Let n be a positive integer and let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces with $\dim(\mathcal{X}) = n^2$, $\dim(\mathcal{Y}) = n$, and $\dim(\mathcal{Z}) = n^2$. Let $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be a linear isometry, and define channels $\Phi, \Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ as*

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(VXV^*\big) \quad \textit{and} \quad \Psi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(\overline{V}XV^{\mathsf{T}}\big) \tag{7.399}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It holds that

$$\mathrm{H}_{\min}\big(\Phi \otimes \Psi\big) \leq 2\log(n) - \frac{\log(n) - 2}{n}. \tag{7.400}$$

Proof. Define pure states $\tau \in \mathrm{D}(\mathcal{X})$ and $\sigma \in \mathrm{D}(\mathcal{Y})$ as follows:

$$\tau = \frac{\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*}{n^2} \quad \text{and} \quad \sigma = \frac{\mathrm{vec}(\mathbb{1}_{\mathcal{Y}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*}{n}. \tag{7.401}$$

A calculation reveals that

$$\big\langle \sigma, (\Phi \otimes \Psi)(\tau) \big\rangle = \frac{1}{n^3} \big\| \mathrm{Tr}_{\mathcal{Y}}(VV^*) \big\|_2^2. \tag{7.402}$$

In greater detail, by selecting $\{y_1, \ldots, y_n\}$ to be any choice of an orthonormal basis of $\mathcal{Y}$, one has

$$\begin{aligned}
&\big\langle \sigma, (\Phi \otimes \Psi)(\tau) \big\rangle \\
&\quad = \frac{1}{n} \sum_{1 \leq j,k \leq n} \Big\langle V^*\big(y_j y_k^* \otimes \mathbb{1}_{\mathcal{Z}}\big)V \otimes V^{\mathsf{T}}\big(\overline{y_j} y_k^{\mathsf{T}} \otimes \mathbb{1}_{\mathcal{Z}}\big)\overline{V}, \tau \Big\rangle \\
&\quad = \frac{1}{n^3} \sum_{1 \leq j,k \leq n} \mathrm{Tr}\Big(\big(V^*\big(y_j y_k^* \otimes \mathbb{1}_{\mathcal{Z}}\big)V\big)\big(V^*\big(y_k y_j^* \otimes \mathbb{1}_{\mathcal{Z}}\big)V\big)\Big) \\
&\quad = \frac{1}{n^3} \big\| \mathrm{Tr}_{\mathcal{Y}}(VV^*) \big\|_2^2.
\end{aligned} \tag{7.403}$$

As the operator $\mathrm{Tr}_{\mathcal{Y}}(VV^*)$ is positive semidefinite, and has trace equal to n^2 and rank at most n^2, it follows that its 2-norm squared must be at least n^2. Consequently, one has

$$\lambda_1\big((\Phi \otimes \Psi)(\tau)\big) \geq \big\langle \sigma, (\Phi \otimes \Psi)(\tau) \big\rangle \geq \frac{1}{n}. \tag{7.404}$$

Now, under the constraint that a given density operator $\rho \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Y})$ has largest eigenvalue at least $1/n$, it holds that the von Neumann entropy

$\mathrm{H}(\rho)$ is maximized when this largest eigenvalue is equal to $1/n$ and all other eigenvalues are equal:

$$\mathrm{H}(\rho) \leq \left(1 - \frac{1}{n}\right) \log(n^2 - 1) + \mathrm{H}\left(\frac{1}{n}, 1 - \frac{1}{n}\right). \tag{7.405}$$

Making use of the bound $\ln(\alpha) \geq 1 - 1/\alpha$, which holds for all positive α, one finds that

$$\mathrm{H}(\lambda, 1-\lambda) \leq -\lambda \log(\lambda) + \frac{\lambda}{\ln(2)} \leq -\lambda \log(\lambda) + 2\lambda \tag{7.406}$$

for all $\lambda \in [0, 1]$, and therefore

$$\mathrm{H}(\rho) \leq 2 \log(n) - \frac{\log(n) - 2}{n}. \tag{7.407}$$

In particular, this inequality holds for $\rho = (\Phi \otimes \Psi)(\tau)$, which completes the proof. □

The remaining lemmas required for the proof of Theorem 7.53 are used to establish a lower bound on the quantity $\mathrm{H}_{\min}(\Phi) + \mathrm{H}_{\min}(\Psi)$, for some choice of channels Φ and Ψ taking the form (7.398). The first lemma is concerned with the modification of a random variable that is Lipschitz on a compact subset of its domain, yielding one that is Lipschitz everywhere.

Lemma 7.55. *Let $\mathcal{X}$ be a complex Euclidean space, let $\mathcal{A} \subseteq \mathcal{S}(\mathcal{X})$ be a non-empty, compact subset of $\mathcal{S}(\mathcal{X})$, and let $X : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$ be a continuous random variable, distributed with respect to the uniform spherical measure μ on $\mathcal{S}(\mathcal{X})$. Let κ be a positive real number such that $|X(x) - X(y)| \leq \kappa \|x - y\|$ for all $x, y \in \mathcal{A}$, and define a new random variable $Y : \mathcal{S}(\mathcal{X}) \to \mathbb{R}$, distributed with respect to μ, as*

$$Y(x) = \min_{y \in \mathcal{A}} \big(X(y) + \kappa \|x - y\| \big) \tag{7.408}$$

for all $x \in \mathcal{S}(\mathcal{X})$. The following statements hold:

1. *Y is κ-Lipschitz.*
2. *For every $x \in \mathcal{A}$, one has that $X(x) = Y(x)$.*
3. *Every median value of Y is a central value of X.*

Proof. The first statement holds regardless of the specific behavior of X on points in $\mathcal{A}$. Consider any two vectors $x_0, x_1 \in \mathcal{S}(\mathcal{X})$, and let $y_0, y_1 \in \mathcal{A}$ safisfy

$$Y(x_0) = X(y_0) + \kappa \|x_0 - y_0\| \quad \text{and} \quad Y(x_1) = X(y_1) + \kappa \|x_1 - y_1\|. \tag{7.409}$$

That is, y_0 and y_1 achieve the minimum values that define the function Y on x_0 and x_1, respectively. It must therefore hold that

$$X(y_0) + \kappa \|x_0 - y_0\| \leq X(y_1) + \kappa \|x_0 - y_1\|, \tag{7.410}$$

which implies

$$Y(x_0) - Y(x_1) \leq \kappa \|x_0 - y_1\| - \kappa \|x_1 - y_1\| \leq \kappa \|x_0 - x_1\|. \tag{7.411}$$

The inequality

$$Y(x_1) - Y(x_0) \leq \kappa \|x_0 - x_1\| \tag{7.412}$$

is proved through the same argument by exchanging the indices 0 and 1. It therefore holds that

$$|Y(x_0) - Y(x_1)| \leq \kappa \|x_0 - x_1\|, \tag{7.413}$$

so Y is κ-Lipschitz.

Next, consider any vector $x \in \mathcal{A}$. By the assumptions of the lemma, one has

$$X(x) - X(y) \leq \kappa \|x - y\| \tag{7.414}$$

for every $y \in \mathcal{A}$, and therefore

$$Y(x) = \min_{y \in \mathcal{A}} \Big(X(y) + \kappa \|x - y\| \Big) \geq X(x). \tag{7.415}$$

On the other hand, because one may choose $y = x$ when considering the minimum, it holds that $Y(x) \leq X(x)$. It follows that $X(x) = Y(x)$, which establishes the second statement.

Finally, let $\alpha \in \mathbb{R}$ be a median value of Y, so that

$$\Pr(Y \geq \alpha) \geq \frac{1}{2} \quad \text{and} \quad \Pr(Y \leq \alpha) \geq \frac{1}{2}. \tag{7.416}$$

Define a random variable $Z : \mathcal{S}(\mathcal{X}) \to [0, 1]$, again distributed with respect to μ, as

$$Z(x) = \begin{cases} 1 & \text{if } x \in \mathcal{A} \\ 0 & \text{if } x \notin \mathcal{A}, \end{cases} \tag{7.417}$$

so that $\Pr(Z = 0) \leq 1/4$. By the union bound, one has

$$\Pr(Y < \alpha \text{ or } Z = 0) \leq \frac{3}{4}, \tag{7.418}$$

and therefore

$$\Pr(X \geq \alpha) \geq \Pr(Y \geq \alpha \text{ and } Z = 1) \geq \frac{1}{4}. \tag{7.419}$$

By similar reasoning,

$$\Pr(X \leq \alpha) \geq \Pr(Y \leq \alpha \text{ and } Z = 1) \geq \frac{1}{4}. \tag{7.420}$$

This implies that α is a central value of X, which completes the proof. □

The next lemma is, in some sense, the heart of the proof of Theorem 7.53. It establishes the existence of an isometry $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ that may be taken in the definition (7.398) of the channels Φ and Ψ to obtain the inequality (7.396) for a sufficiently large value of n. It is proved through the use of Dvoretzky's theorem.

Lemma 7.56. *There exists a real number $K > 0$ for which the following statement holds. For every choice of a positive integer n, and for $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces with*

$$\dim(\mathcal{X}) = n^2, \quad \dim(\mathcal{Y}) = n, \quad \text{and} \quad \dim(\mathcal{Z}) = n^2, \tag{7.421}$$

there exists an isometry $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ such that

$$\left\| \mathrm{Tr}_{\mathcal{Z}}(Vxx^*V^*) - \omega \right\|_2 \leq \frac{K}{n} \tag{7.422}$$

for every unit vector $x \in \mathcal{S}(\mathcal{X})$, where $\omega = \mathbb{1}/n$ denotes the completely mixed state with respect to $\mathcal{Y}$.

Proof. Let δ be a positive real number that satisfies the requirements of Dvoretzky's theorem (Theorem 7.46) for $\zeta = 1/3$, and let K_0 be a positive real number satisfying the requirements of Lemma 7.49. It will be proved that the lemma holds for

$$K = K_0 + 2\sqrt{\frac{K_0 + 1}{\delta}} + \frac{2}{\delta}. \tag{7.423}$$

For the remainder of the proof it will be assumed that a positive integer n and complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ satisfying (7.421) have been fixed. Let $\mathcal{V}$ be an arbitrary subspace of $\mathcal{Y} \otimes \mathcal{Z}$ having dimension n^2. Throughout the proof, μ will denote the uniform spherical measure on $\mathcal{S}(\mathcal{Y} \otimes \mathcal{Z})$, and η will denote the Haar measure on $\mathrm{U}(\mathcal{Y} \otimes \mathcal{Z})$.

The first step of the proof is the specification of a collection of random variables; an analysis of these random variables follows their specification. First, let $X, Y : \mathcal{S}(\mathcal{Y} \otimes \mathcal{Z}) \to \mathbb{R}$ be random variables, distributed with respect to the uniform spherical measure μ and defined as follows:

$$X(u) = \sqrt{\left\| \mathrm{Tr}_{\mathcal{Z}}(uu^*) \right\|} \quad \text{and} \quad Y(u) = \left\| \mathrm{Tr}_{\mathcal{Z}}(uu^*) - \omega \right\|_2 \tag{7.424}$$

for all $u \in \mathcal{S}(\mathcal{Y} \otimes \mathcal{Z})$. Next, let

$$K_1 = \sqrt{K_0 + 1} + \frac{1}{\sqrt{\delta}} \quad \text{and} \quad \kappa = \frac{2K_1}{\sqrt{n}}, \tag{7.425}$$

define a set

$$\mathcal{A} = \left\{ u \in \mathcal{S}(\mathcal{Y} \otimes \mathcal{Z}) \,:\, X(u) \leq \frac{K_1}{\sqrt{n}} \right\}, \tag{7.426}$$

and define a random variable $Z : \mathcal{S}(\mathcal{Y} \otimes \mathcal{Z}) \to \mathbb{R}$, also distributed with respect to the uniform spherical measure μ, as

$$Z(u) = \min_{v \in \mathcal{A}} \big(Y(v) + \kappa \| u - v \| \big) \tag{7.427}$$

for every $u \in \mathcal{S}(\mathcal{Y} \otimes \mathcal{Z})$. It is evident from their specifications that X, Y, and Z are phase-invariant random variables. Finally, for each unit vector $v \in \mathcal{S}(\mathcal{V})$, define random variables $P_v, Q_v, R_v : \mathrm{U}(\mathcal{Y} \otimes \mathcal{Z}) \to \mathbb{R}$, distributed with respect to the Haar measure η on $\mathrm{U}(\mathcal{Y} \otimes \mathcal{Z})$, as

$$P_v(U) = X(Uv), \quad Q_v(U) = Y(Uv), \quad \text{and} \quad R_v(U) = Z(Uv), \tag{7.428}$$

for every $U \in \mathrm{U}(\mathcal{Y} \otimes \mathcal{Z})$.

When analyzing the random variables that have just been defined, it is helpful to begin with the observation that

$$X(\mathrm{vec}(A)) = \| A \| \quad \text{and} \quad Y(\mathrm{vec}(A)) = \left\| AA^* - \omega \right\|_2 \tag{7.429}$$

for every operator $A \in \mathrm{L}(\mathcal{Z}, \mathcal{Y})$ satisfying $\| A \|_2 = 1$. It is immediate from the first of these expressions, along with the inequality $\| A \| \leq \| A \|_2$, that X is 1-Lipschitz. Also, given that

$$\| A \|^2 = \left\| AA^* \right\| \leq \left\| AA^* - \omega \right\| + \| \omega \| \leq \left\| AA^* - \omega \right\|_2 + \frac{1}{n} \tag{7.430}$$

for every operator $A \in \mathrm{L}(\mathcal{Z}, \mathcal{Y})$, one necessarily has that

$$X^2 \leq Y + \frac{1}{n}. \tag{7.431}$$

By Lemma 7.49, one may therefore conclude that

$$\Pr\left(X \leq \sqrt{\frac{K_0 + 1}{n}}\right) \geq \Pr\left(Y \leq \frac{K_0}{n}\right) > \frac{3}{4}. \tag{7.432}$$

Dvoretzky's theorem (Theorem 7.46) will be applied twice in the proof, with the first application concerning the random variables X and P_v for each $v \in \mathcal{S}(\mathcal{V})$. By (7.432), it follows that every central value of X is at most

$$\sqrt{\frac{K_0 + 1}{n}}. \tag{7.433}$$

Setting

$$\varepsilon = \frac{1}{\sqrt{\delta n}} \tag{7.434}$$

in Dvoretzky's theorem yields

$$\Pr\left(P_v \leq \frac{K_1}{\sqrt{n}} \text{ for every } v \in \mathcal{S}(\mathcal{V})\right) \geq \frac{2}{3}, \tag{7.435}$$

by virtue of the fact that $\dim(\mathcal{V}) = \delta\varepsilon^2 \dim(\mathcal{Y} \otimes \mathcal{Z})$.

The second application of Dvoretzky's theorem concerns the random variables Z and R_v for each $v \in \mathcal{S}(\mathcal{V})$. Before applying Dvoretzky's theorem, however, the implications of Lemma 7.55 to the random variables Y and Z will be considered. First, note that

$$\mu(\mathcal{A}) = \Pr\left(X \leq \frac{K_1}{\sqrt{n}}\right) \geq \Pr\left(X \leq \sqrt{\frac{K_0 + 1}{n}}\right) > \frac{3}{4}. \tag{7.436}$$

Second, for any choice of vectors $u, v \in \mathcal{A}$, one may write $u = \mathrm{vec}(A)$ and $v = \mathrm{vec}(B)$ for $A, B \in \mathrm{L}(\mathcal{Z}, \mathcal{Y})$ satisfying $\|A\|_2 = \|B\|_2 = 1$, so that

$$\|A\| = X(\mathrm{vec}(A)) \leq \frac{K_1}{\sqrt{n}} \quad \text{and} \quad \|B\| = X(\mathrm{vec}(B)) \leq \frac{K_1}{\sqrt{n}}. \tag{7.437}$$

This implies that

$$\begin{aligned}\big|Y(u) - Y(v)\big| &= \Big|\big\|AA^* - \omega\big\|_2 - \big\|BB^* - \omega\big\|_2\Big| \\ &\leq \big\|AA^* - BB^*\big\|_2 \leq \big(\|A\| + \|B\|\big)\big\|A - B\big\|_2 \leq \kappa\|u - v\|.\end{aligned} \tag{7.438}$$

It therefore follows from Lemma 7.55 that Z is κ-Lipschitz, Z and Y agree everywhere on $\mathcal{A}$, and every median value of Z is a central value of Y. By (7.432), every central value of Y is at most K_0/n, and therefore the same upper bound applies to every median value of Z. Setting

$$\varepsilon = \frac{\kappa}{\sqrt{\delta n}} \tag{7.439}$$

and applying Dvoretzky's theorem therefore yields

$$\Pr\left(R_v \leq \frac{K}{n} \text{ for all } v \in \mathcal{S}(\mathcal{V})\right) \geq \frac{2}{3}, \tag{7.440}$$

by virtue of the fact that

$$\dim(\mathcal{V}) = \frac{\delta\varepsilon^2}{\kappa^2}\dim(\mathcal{Y} \otimes \mathcal{Z}). \tag{7.441}$$

Finally, consider the random variables Y and Q_v for each $v \in \mathcal{S}(\mathcal{V})$. For every vector $u \in \mathcal{S}(\mathcal{Y} \otimes \mathcal{Z})$, one has either $u \in \mathcal{A}$ or $u \notin \mathcal{A}$; and if it holds that $u \in \mathcal{A}$, then $Y(u) = Z(u)$. Consequently, if it holds that $Y(u) > K/n$ for a given choice of $u \in \mathcal{S}(\mathcal{Y} \otimes \mathcal{Z})$, then it must hold that

$$Z(u) > \frac{K}{n} \quad \text{or} \quad X(u) > \frac{K_1}{\sqrt{n}} \tag{7.442}$$

(or both). By the union bound, one concludes that

$$\begin{aligned}&\Pr\left(Q_v > \frac{K}{n} \text{ for some } v \in \mathcal{S}(\mathcal{V})\right) \\ &\qquad \leq \Pr\left(R_v > \frac{K}{n} \text{ for some } v \in \mathcal{S}(\mathcal{V})\right) \\ &\qquad\quad + \Pr\left(P_v > \frac{K_1}{\sqrt{n}} \text{ for some } v \in \mathcal{S}(\mathcal{V})\right).\end{aligned} \tag{7.443}$$

By (7.435) and (7.440), it follows that

$$\Pr\left(Q_v \leq \frac{K}{n} \text{ for all } v \in \mathcal{S}(\mathcal{V})\right) \geq \frac{1}{3} > 0. \tag{7.444}$$

By (7.444), one concludes that there exists a unitary operator U for which $Q_v(U) \leq K/n$ for all $v \in \mathcal{S}(\mathcal{V})$. Taking $V_0 \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ to be any linear isometry for which $\mathrm{im}(V_0) = \mathcal{V}$, one therefore has

$$\left\| \mathrm{Tr}_{\mathcal{Z}}\left(UV_0 x x^* V_0^* U^*\right) - \omega \right\|_2 \leq \frac{K}{n} \tag{7.445}$$

for every unit vector $x \in \mathcal{S}(\mathcal{X})$. Taking $V = UV_0$, the lemma is proved. □

Finally, a proof of Theorem 7.53 is to be presented. The proof is made quite straightforward through the use of Lemmas 7.54 and 7.56.

Proof of Theorem 7.53. Let $K > 0$ be a real number for which Lemma 7.56 holds, and choose n to be a positive integer satisfying

$$\log(n) > \frac{2K^2}{\ln(2)} + 2. \tag{7.446}$$

For $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces with $\dim(\mathcal{X}) = n^2$, $\dim(\mathcal{Y}) = n$, and $\dim(\mathcal{Z}) = n^2$, it follows (by Lemma 7.56) that there exists an isometry $V \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ such that

$$\left\| \mathrm{Tr}_{\mathcal{Z}}\left(V x x^* V^*\right) - \frac{\mathbb{1}_{\mathcal{Y}}}{n} \right\|_2 \leq \frac{K}{n} \tag{7.447}$$

for every unit vector $x \in \mathcal{S}(\mathcal{X})$. By Lemma 7.50, one therefore has that

$$\mathrm{H}\big(\mathrm{Tr}_{\mathcal{Z}}(V x x^* V^*)\big) \geq \log(n) - \frac{K^2}{n \ln(2)} \tag{7.448}$$

for every $x \in \mathcal{S}(\mathcal{X})$. Replacing V by the entrywise complex conjugate of V results in the same bound:

$$\mathrm{H}\big(\mathrm{Tr}_{\mathcal{Z}}\big(\overline{V} x x^* V^{\mathsf{T}}\big)\big) \geq \log(n) - \frac{K^2}{n \ln(2)} \tag{7.449}$$

for every $x \in \mathcal{S}(\mathcal{X})$.

Now, define channels $\Phi, \Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ as

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\left(VXV^*\right) \quad \text{and} \quad \Psi(X) = \mathrm{Tr}_{\mathcal{Z}}\left(\overline{V} X V^{\mathsf{T}}\right) \tag{7.450}$$

for all $X \in \mathrm{L}(\mathcal{X})$. One has that

$$\mathrm{H}_{\min}(\Phi) = \mathrm{H}_{\min}(\Psi) \geq \log(n) - \frac{K^2}{n \ln(2)}, \tag{7.451}$$

and therefore

$$\mathrm{H}_{\min}(\Phi) + \mathrm{H}_{\min}(\Psi) \geq 2\log(n) - \frac{2K^2}{n\ln(2)}. \tag{7.452}$$

On the other hand, Lemma 7.54 implies that

$$\mathrm{H}_{\min}(\Phi \otimes \Psi) \leq 2\log(n) - \frac{\log(n) - 2}{n}. \tag{7.453}$$

Consequently,

$$\begin{aligned} &\mathrm{H}_{\min}(\Phi \otimes \Psi) - \big(\mathrm{H}_{\min}(\Phi) + \mathrm{H}_{\min}(\Psi)\big) \\ &\qquad = \frac{2K^2}{n\ln(2)} - \frac{\log(n) - 2}{n} < 0, \end{aligned} \tag{7.454}$$

which completes the proof. □

7.4 Exercises

7.1. For every positive integer $n \geq 2$, define a unital channel $\Phi_n \in \mathrm{C}(\mathbb{C}^n)$ as

$$\Phi_n(X) = \frac{1}{n-1}\operatorname{Tr}(X)\mathbb{1}_n - \frac{1}{n-1}X^{\mathsf{T}} \tag{7.455}$$

for every $X \in \mathrm{L}(\mathbb{C}^n)$, where $\mathbb{1}_n$ denotes the identity operator on $\mathbb{C}^n$. Prove that Φ_n is a mixed-unitary channel when n is even.

Observe that this exercise is complementary to Exercise 4.2.

7.2. Let n and m be positive integers with $n < m$, and consider the set $\mathrm{U}(\mathbb{C}^n, \mathbb{C}^m)$ of all isometries from $\mathbb{C}^n$ to $\mathbb{C}^m$.

(a) Prove that there exists a probability measure

$$\nu : \mathrm{Borel}\big(\mathrm{U}(\mathbb{C}^n, \mathbb{C}^m)\big) \to [0, 1] \tag{7.456}$$

for which it holds that

$$\nu(\mathcal{A}) = \nu(U\mathcal{A}V) \tag{7.457}$$

for every choice of unitary operators $U \in \mathrm{U}(\mathbb{C}^m)$ and $V \in \mathrm{U}(\mathbb{C}^n)$.

(b) Prove that if

$$\mu : \text{Borel}\big(\mathrm{U}(\mathbb{C}^n, \mathbb{C}^m)\big) \to [0,1] \tag{7.458}$$

is a probability measure satisfying

$$\mu(\mathcal{A}) = \mu(U\mathcal{A}) \tag{7.459}$$

for every unitary operator $U \in \mathrm{U}(\mathbb{C}^m)$, then it must hold that $\mu = \nu$, where ν is the measure defined by a correct solution to part (a).

7.3. Let $\mathcal{X}$ be a complex Euclidean space, let $n = \dim(\mathcal{X})$, and define a mapping $\Phi \in \mathrm{CP}(\mathcal{X})$ as

$$\Phi(X) = n \int \big\langle uu^*, X \big\rangle uu^* \, \mathrm{d}\mu(u) \tag{7.460}$$

for all $X \in \mathrm{L}(\mathcal{X})$, where μ denotes the uniform spherical measure on $\mathcal{S}(\mathcal{X})$. Give a simple, closed-form expression for Φ.

7.4. Let $\mathcal{X}$ be a complex Euclidean space, let $n = \dim(\mathcal{X})$, and define a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{X} \otimes \mathcal{X})$ as

$$\Phi(X) = n \int \big\langle uu^*, X \big\rangle uu^* \otimes uu^* \, \mathrm{d}\mu(u) \tag{7.461}$$

for all $X \in \mathrm{L}(\mathcal{X})$, where μ denotes the uniform spherical measure on $\mathcal{S}(\mathcal{X})$. Give a closed-form expression for the minimum cloning fidelity

$$\alpha(\Phi) = \inf_{v \in \mathcal{S}(\mathcal{X})} \mathrm{F}\big(\Phi(vv^*), vv^* \otimes vv^*\big) \tag{7.462}$$

obtained through the use of Φ. Observe that Φ is a sub-optimal cloning channel, in the sense of Theorem 7.29 (aside from the trivial case in which $\dim(\mathcal{X}) = 1$).

7.5. Prove that there exists a real number $K > 0$ with the following property. For every integer $n \geq 1$ and every nonnegative κ-Lipschitz random variable

$$X : \mathcal{S}(\mathbb{C}^n) \to [0, \infty), \tag{7.463}$$

distributed with respect to the uniform spherical measure on $\mathcal{S}(\mathbb{C}^n)$, one has that

$$\mathrm{E}(X^2) - \mathrm{E}(X)^2 \leq \frac{K\kappa^2}{n}. \tag{7.464}$$

7.6. Prove that there exist positive real numbers $K, \delta > 0$ for which the following statement holds. For every choice of a complex Euclidean space $\mathcal{X}$, a κ-Lipschitz nonnegative random variable

$$X : \mathcal{S}(\mathcal{X}) \to [0, \infty), \tag{7.465}$$

distributed with respect to the uniform spherical measure μ on $\mathcal{S}(\mathcal{X})$, and every positive real number $\varepsilon > 0$, it holds that

$$\Pr\left(\left|X - \sqrt{\mathrm{E}(X^2)}\right| \geq \varepsilon\right) \leq K \exp\left(-\frac{\delta \varepsilon^2 n}{\kappa^2}\right). \tag{7.466}$$

The fact established by a correct solution to Exercise 7.5 is useful for proving this result.

Observe that a correct solution to this problem etablishes a variant of Lévy's lemma in which concentration occurs around the root-mean-squared value of a nonnegative random variable, as opposed to its mean or central values.

7.5 Bibliographic remarks

Permutation-invariant vectors and operators are commonly studied objects in multilinear algebra, which is the subject of the books of Greub [81] and Marcus [152, 153], among others. These concepts and generalizations of them are also relevant to the subject of representation theory, as explained in the book of Goodman and Wallach [79], for instance. Theorem 7.15 is a finite-dimensional form of the double commutant theorem, also known as the bicommutant theorem, proved by von Neumann in 1930 [215].

The existence of unitarily invariant measures on both the unit sphere and the set of unitary operators in a complex Euclidean space is implied by a much more general construction due to Haar [87]. Von Neumann [216] proved the uniqueness of the measures constructed by Haar, with their two papers appearing consecutively in the same journal. This work was further generalized by Weil [224] and others. Due to the generality of these notions, many books that include a discussion of Haar measure do not consider the specialized definitions of uniform spherical measure or Haar measure (for unitary operators in finite dimensions) of the sort that has been presented in this chapter. Definitions of this type are, however, fairly standard in random

matrix theory. These definitions are rooted in the work of Dyson [64, 65, 66] and Diaconis and Shahshahani [60], and a more broad overview of random matrix theory may be found in the book of Mehta [156].

The Werner twirling channel, defined in Example 7.26, was introduced by Werner [225] in the same paper, mentioned in the previous chapter, that introduced the states now known as Werner states. Theorem 7.29 on optimal cloning of pure states is also due to Werner [226].

Multiple versions of the quantum de Finetti theorem are known. These theorems are so-named because they generalize theorems in combinatorics and probability theory originally found in the work of de Finetti [56]. A quantum information-theoretic variant of de Finetti's eponymous theorem was first proved by Hudson and Moody [122] in 1976. Caves, Fuchs, and Schack [44] later gave a simpler proof of this theorem. Like the original de Finetti theorem, this was a qualitative result regarding the behavior of an infinite number of identical systems. A finite quantum formulation of de Finetti's theorem, closer in spirit to classical results due to Diaconis and Freedman [59], was proved by König and Renner [133]. Theorems 7.13 and 7.27 and Corollary 7.28 were proved by Christandl, König, Mitchison, and Renner [51], who improved on the error bounds and generalized the results obtained by König and Renner.

Theorem 7.32 and Corollary 7.33 are due to Watrous [221].

Readers interested in learning more about the phenomenon of measure concentration are referred to the books of Ledoux [144] and Milman and Schechtman [158]. Theorems 7.39 and 7.43 are variants of a theorem due to Lévy [145]. The proofs of these theorems appearing in this chapter have mostly followed those in Appendix V of Milman and Schechtman's book (which are partially based on a technique due to Maurey and Pisier [155]). Multiple formulations of Dvoretsky's theorem are known, with the original having been proved by Dvoretsky around 1960 [63]. Milman [157] gave a proof of Dvoretzky's theorem in 1971 based on the measure concentration phenomenon, which he was the first to explicitly identify.

To prove Theorem 7.53 on the non-additivity of the minimum output entropy, a particularly sharp version of Dvoretzky's theorem (as stated in Theorem 7.46) is evidently required. The proof of this theorem, as well as its application to Theorem 7.53, is due to Aubrun, Szarek, and Werner [16]. The proof makes essential use of Talagrand's *chaining method* [199].

There are several known applications of the concentration of measure

phenomenon to quantum information theory, the first of which were due to Hayden, Leung, Shor, and Winter [96], Bennett, Hayden, Leung, Shor, and Winter [37], and Harrow, Hayden, and Leung [90]. Theorem 7.51 is a variant of a theorem due to Hayden, Leung, and Winter [97]. Theorem 7.53 was proved by Hastings [91], based in part on Hayden and Winter's disproof of the so-called *maximal p-norm multiplicativity conjecture* shortly before [99]. As suggested above, the proof of Theorem 7.53 that has been presented in this chapter is due to Aubrun, Szarek, and Werner [16]. The implications of Hastings discovery to the study of channel capacities is discussed in the next chapter.

Chapter 8

Quantum channel capacities

This chapter is focused on *capacities* of quantum channels for transmitting information. The notion of a channel capacity has multiple, inequivalent formulations in the quantum setting. For example, one may consider the capacity with which classical or quantum information can be transmitted through a channel, and different resources may or may not be available to assist the information transmission—such as entanglement shared between a sender and receiver before the information transmission takes place.

Three fundamental theorems are presented, characterizing the capacities of quantum channels to transmit either classical or quantum information, both with and without the assistance of prior shared entanglement. When prior shared entanglement between the sender and receiver is not available, these characterizations have a somewhat undesirable property: they require a *regularization*—or an averaging over an increasingly large number of uses of a given channel—and fail to provide capacity formulas that are either explicit or efficiently computable for this reason. The apparent need for such regularizations is discussed in the last section of the chapter, along with the related phenomenon of *super-activation* of quantum capacity.

8.1 Classical information over quantum channels

The general scenario to be considered throughout this chapter involves two hypothetical individuals: a *sender* and a *receiver*. The sender wishes to transmit information, either classical or quantum, to the receiver, and is able to do this through multiple, independent uses of a given channel Φ. One aims

to design a scheme by which the sender prepares an input to these channel uses and the receiver processes their output in such a way that information is transmitted with a high degree of accuracy. As is standard in information theory, the chapter mainly deals with the asymptotic regime, making use of entropic notions to analyze rates of information transmission in the limit of an increasingly large number of independent channel uses.

The subject of the present section is the capacity of quantum channels to transmit *classical* information, including both the case in which the sender and receiver share prior entanglement and in which they do not. The first subsection below introduces notions and terminology concerning channel capacities that will be needed throughout the section, as well as in later parts of the chapter. The second subsection is devoted to a proof of the *Holevo–Schumacher–Westmoreland theorem*, which characterizes the capacity of a channel to transmit classical information without the use of prior shared entanglement. The final subsection proves the *entanglement-assisted capacity theorem*, which characterizes the capacity of a channel to transmit classical information with the assistance of prior shared entanglement.

8.1.1 Classical capacities of quantum channels

Five quantities that relate to the information-transmitting capabilities of channels are defined below. The first two quantities are fundamental with respect to the subject of quantum channel capacities: the *classical capacity* and the *entanglement-assisted classical capacity* of a quantum channel. The remaining three quantities are the *Holevo capacity*, the *entanglement-assisted Holevo capacity*, and the *coherent information*, all of which play important roles in the main results to be presented.

The classical capacity of a channel

Intuitively (and somewhat informally) speaking, the classical capacity of a channel describes 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.

When stating a precise mathematical definition of classical capacity, it is convenient to refer to the *emulation* of one channel by another, as well as to

the *approximation* of one channel by another, with the approximation being defined with respect to the completely bounded trace norm.

Definition 8.1. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{Z})$ be channels, for $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces. It is said that the channel Φ *emulates* Ψ if there exist channels $\Xi_{\mathrm{E}} \in \mathrm{C}(\mathcal{Z}, \mathcal{X})$ and $\Xi_{\mathrm{D}} \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ such that

$$\Psi = \Xi_{\mathrm{D}} \Phi \, \Xi_{\mathrm{E}}. \tag{8.1}$$

When this relationship holds, the channel Ξ_{E} is called an *encoding channel* and Ξ_{D} is called a *decoding channel*.

Definition 8.2. Let $\mathcal{Z}$ be a complex Euclidean space, let $\Psi_0, \Psi_1 \in \mathrm{C}(\mathcal{Z})$ be channels, and let $\varepsilon > 0$ be a positive real number. The channel Ψ_0 is an ε-approximation to Ψ_1 (or, equivalently, Ψ_1 is an ε-approximation to Ψ_0) if and only if

$$\bigl|\!\bigl|\!\bigl| \Psi_0 - \Psi_1 \bigr|\!\bigr|\!\bigr|_1 < \varepsilon. \tag{8.2}$$

The definition of the classical capacity of a quantum channel, which makes use of the previous two definitions, is as follows.

Definition 8.3 (Classical capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Let $\Gamma = \{0, 1\}$ denote the binary alphabet, let $\mathcal{Z} = \mathbb{C}^{\Gamma}$, and let $\Delta \in \mathrm{C}(\mathcal{Z})$ denote the completely dephasing channel defined with respect to the space $\mathcal{Z}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for classical information transmission through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the channel $\Delta^{\otimes m}$.
2. The *classical capacity* of Φ, denoted $\mathrm{C}(\Phi)$, is the supremum value of all achievable rates for classical information transmission through Φ.

In the context of Definition 8.3, the completely dephasing channel Δ is to be viewed as an ideal channel for transmitting a single bit of classical information. When considering an emulation of the m-fold tensor product $\Delta^{\otimes m}$ of this ideal classical channel by the channel $\Phi^{\otimes n}$, no generality is lost in restricting one's attention to classical-to-quantum encoding channels Ξ_{E}

and quantum-to-classical decoding channels Ξ_{D}. That is, one may assume that the conditions

$$\Xi_{\mathrm{E}} = \Xi_{\mathrm{E}}\Delta^{\otimes m} \quad \text{and} \quad \Xi_{\mathrm{D}} = \Delta^{\otimes m}\Xi_{\mathrm{D}} \tag{8.3}$$

are met. This is so because

$$\begin{aligned} &\bigl|\!\bigl|\!\bigl| \bigl(\Delta^{\otimes m}\Xi_{\mathrm{D}}\bigr)\Phi^{\otimes n}\bigl(\Xi_{\mathrm{E}}\Delta^{\otimes m}\bigr) - \Delta^{\otimes m} \bigr|\!\bigr|\!\bigr|_1 \\ &\qquad = \bigl|\!\bigl|\!\bigl| \Delta^{\otimes m}\bigl(\Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} - \Delta^{\otimes m}\bigr)\Delta^{\otimes m} \bigr|\!\bigr|\!\bigr|_1 \\ &\qquad \leq \bigl|\!\bigl|\!\bigl| \Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} - \Delta^{\otimes m} \bigr|\!\bigr|\!\bigr|_1, \end{aligned} \tag{8.4}$$

which implies that replacing a given choice of Ξ_{E} and Ξ_{D} by $\Xi_{\mathrm{E}}\Delta^{\otimes m}$ and $\Delta^{\otimes m}\Xi_{\mathrm{D}}$ cannot decrease the quality of the emulation that is achieved.

In light of this observation, the implicit use of the completely bounded trace norm in Definition 8.3 may appear to be somewhat heavy-handed; an equivalent definition is obtained by requiring that $\Phi^{\otimes n}$ emulates any channel $\Psi \in \mathrm{C}(\mathcal{Z}^{\otimes m})$ satisfying

$$\bigl\| \bigl(\Delta^{\otimes m}\Psi\bigr)\bigl(E_{a_1\cdots a_m,a_1\cdots a_m}\bigr) - E_{a_1\cdots a_m,a_1\cdots a_m} \bigr\|_1 < \varepsilon, \tag{8.5}$$

which is equivalent to

$$\bigl\langle E_{a_1\cdots a_m,a_1\cdots a_m}, \Psi\bigl(E_{a_1\cdots a_m,a_1\cdots a_m}\bigr)\bigr\rangle > 1 - \frac{\varepsilon}{2}, \tag{8.6}$$

for all $a_1\cdots a_m \in \Gamma^m$. An interpretation of this requirement is that every string $a_1\cdots a_m \in \Gamma^m$ is transmitted by Ψ with a probability of error smaller than $\varepsilon/2$.

There is, on the other hand, a benefit to using the more general notion of channel approximation defined by the completely bounded trace norm in Definition 8.3, which is that it allows the quantum capacity to be defined in an analogous manner to the classical capacity—replacing the dephasing channel Δ by the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}$. (For the quantum capacity, which is discussed later in Section 8.2, the completely bounded trace norm provides a natural notion of channel approximation.)

The following proposition is, perhaps, self-evident, but it is nevertheless worth stating explicitly. The same argument used to prove it may be applied to other notions of capacity as well—there is nothing specific to the classical capacity that is required by the proof.

Proposition 8.4. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let k be a positive integer. It holds that $\mathrm{C}(\Phi^{\otimes k}) = k\,\mathrm{C}(\Phi)$.*

Proof. Assume first that $\alpha \geq 0$ is an achievable rate for classical information transmission through Φ. It follows immediately that αk is an achievable rate for information transmission through $\Phi^{\otimes k}$, and therefore

$$\mathrm{C}\big(\Phi^{\otimes k}\big) \geq k\,\mathrm{C}(\Phi). \tag{8.7}$$

Next, assume that α is an achievable rate for information transmission through $\Phi^{\otimes k}$. If it is the case that $\alpha = 0$, then α/k is trivially an achievable rate for classical information transmission through Φ, so one may focus on the case that $\alpha > 0$. For any choice of a positive integer $n \geq k$, the channel $\Phi^{\otimes n}$ evidently emulates every channel emulated by the channel

$$\Phi^{\otimes k\lfloor n/k \rfloor}. \tag{8.8}$$

For every choice of $\varepsilon > 0$, all but finitely many positive integers n, and all positive integers $m \leq \alpha\lfloor n/k \rfloor$, the channel $\Phi^{\otimes n}$ therefore emulates an ε-approximation to $\Delta^{\otimes m}$. For all $\delta \in (0, \alpha/k)$, it holds that

$$\alpha\lfloor n/k \rfloor \geq (\alpha/k - \delta)n \tag{8.9}$$

for all but finitely many positive integers n, implying that $\alpha/k - \delta$ is an achievable rate for classical information transmission through Φ.

Taking the supremum over all achievable rates, one finds that

$$\mathrm{C}(\Phi) \geq \frac{1}{k}\,\mathrm{C}\big(\Phi^{\otimes k}\big), \tag{8.10}$$

which completes the proof. □

The entanglement-assisted classical capacity of a channel

The entanglement-assisted classical capacity of a channel is defined in a similar way to the classical capacity, except that one assumes that the sender and receiver may share any entangled state of their choosing prior to the transmission of information through the channel. The ability of the sender and receiver to share entanglement, as compared with the situation in which they do not, can result in a significant increase in the classical capacity of a

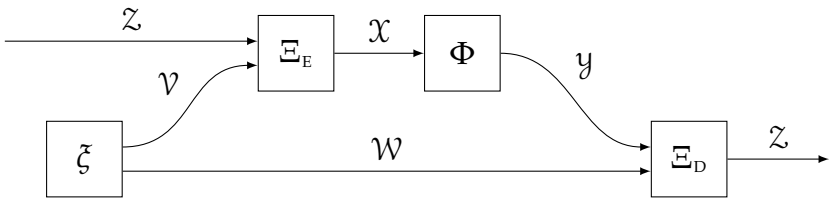


Figure 8.1: An illustration of the channel $\Psi(Z) = \big(\Xi_D\big(\Phi\Xi_E \otimes \mathbb{1}_{L(\mathcal{W})}\big)\big)(Z \otimes \xi)$ referred to in Definition 8.5.

quantum channel. For instance, shared entanglement doubles the classical capacity of the identity channel through the use of dense coding (discussed in Section 6.3.1), and an arbitrary (constant-factor) increase is possible for other choices of channels.

A formal definition for the entanglement-assisted classical capacity of a channel requires only a minor change to the definition of the ordinary classical capacity. In particular, the definition of an emulation of one channel by another is modified to allow for the existence of a shared entangled state as follows.

Definition 8.5. Let $\Phi \in C(\mathcal{X}, \mathcal{Y})$ and $\Psi \in C(\mathcal{Z})$ be channels, for $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ being complex Euclidean spaces. It is said that the channel Φ *emulates* Ψ *with the assistance of entanglement* if and only if there exist complex Euclidean spaces $\mathcal{V}$ and $\mathcal{W}$, a state $\xi \in D(\mathcal{V} \otimes \mathcal{W})$, and channels $\Xi_E \in C(\mathcal{Z} \otimes \mathcal{V}, \mathcal{X})$ and $\Xi_D \in C(\mathcal{Y} \otimes \mathcal{W}, \mathcal{Z})$ such that

$$\Psi(Z) = \big(\Xi_D\big(\Phi\Xi_E \otimes \mathbb{1}_{L(\mathcal{W})}\big)\big)(Z \otimes \xi) \tag{8.11}$$

for all $Z \in L(\mathcal{Z})$. (See Figure 8.1 for an illustration of a channel Ψ satisfying this equation for all $Z \in L(\mathcal{Z})$.) When this relationship holds, the channel Ξ_E is called an *encoding channel*, Ξ_D is called a *decoding channel*, and ξ is referred to as the *shared entangled state* that assists this emulation.

Aside from the modification represented by the previous definition, the entanglement-assisted classical capacity is defined in an analogous way to the ordinary classical capacity.

Definition 8.6 (Entanglement-assisted classical capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in C(\mathcal{X}, \mathcal{Y})$ be a channel. Let $\Gamma = \{0, 1\}$ denote the binary alphabet, let $\mathcal{Z} = \mathbb{C}^\Gamma$, and let $\Delta \in C(\mathcal{Z})$ denote the completely dephasing channel defined with respect to the space $\mathcal{Z}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for entanglement-assisted classical information transmission through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the channel $\Delta^{\otimes m}$ with the assistance of entanglement.
2. The *entanglement-assisted classical capacity* of Φ, which is denoted $\mathrm{C}_{\mathrm{E}}(\Phi)$, is defined as the supremum over all achievable rates for entanglement-assisted classical information transmission through Φ.

Through the same argument used to prove Proposition 8.4, one has that the following simple proposition holds.

Proposition 8.7. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let k be a positive integer. It holds that $\mathrm{C}_{\mathrm{E}}(\Phi^{\otimes k}) = k\,\mathrm{C}_{\mathrm{E}}(\Phi)$.*

The Holevo capacity of a channel

Suppose that $\mathcal{X}$ is a complex Euclidean space, Σ is an alphabet, $p \in \mathcal{P}(\Sigma)$ is a probability vector, and $\{\rho_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X})$ is a collection of states. The Holevo information $\chi(\eta)$ of the ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ defined as

$$\eta(a) = p(a)\rho_a \tag{8.12}$$

for each $a \in \Sigma$ is given by

$$\chi(\eta) = \mathrm{H}\left(\sum_{a\in\Sigma} p(a)\rho_a\right) - \sum_{a\in\Sigma} p(a)\,\mathrm{H}(\rho_a). \tag{8.13}$$

Based on this quantity, one may define the *Holevo capacity* of a channel in the manner specified by Definition 8.8 below. This definition will make use of the following notation: for any ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ and any channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, one defines the ensemble $\Phi(\eta) : \Sigma \to \mathrm{Pos}(\mathcal{Y})$ as

$$(\Phi(\eta))(a) = \Phi(\eta(a)) \tag{8.14}$$

for each $a \in \Sigma$. That is, $\Phi(\eta)$ is the ensemble obtained by evaluating Φ on the ensemble η in the most natural way.

Definition 8.8. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. The *Holevo capacity* of Φ is defined as

$$\chi(\Phi) = \sup_{\eta} \chi(\Phi(\eta)), \tag{8.15}$$

where the supremum is over all choices of an alphabet Σ and an ensemble of the form $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$.

Two restrictions may be placed on the supremum (8.15) in Definition 8.8 without decreasing the value that is defined for a given channel. The first restriction is that the supremum may be replaced by a maximum over all ensembles of the form $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$, for Σ being an alphabet of size

$$|\Sigma| = \dim(\mathcal{X})^2. \tag{8.16}$$

The second restriction is that one may restrict their attention to ensembles η for which $\mathrm{rank}(\eta(a)) \leq 1$ for each $a \in \Sigma$. The following proposition will be used in the proof that this is so.

Proposition 8.9. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let Σ be an alphabet, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble. There exists an alphabet Γ and an ensemble $\tau : \Gamma \to \mathrm{Pos}(\mathcal{X})$ such that*

1. $\mathrm{rank}(\tau(b)) \leq 1$ *for each* $b \in \Gamma$, *and*
2. $\chi(\Phi(\eta)) \leq \chi(\Phi(\tau))$.

Proof. Assume that Δ is the alphabet for which $\mathcal{X} = \mathbb{C}^{\Delta}$, and let

$$\eta(a) = \sum_{c \in \Delta} \lambda_{a,c} x_{a,c} x_{a,c}^* \tag{8.17}$$

be a spectral decomposition of $\eta(a)$ for each $a \in \Sigma$. The requirements of the proposition hold for the ensemble $\tau : \Sigma \times \Delta \to \mathrm{Pos}(\mathcal{X})$ defined by

$$\tau(a, c) = \lambda_{a,c} x_{a,c} x_{a,c}^* \tag{8.18}$$

for each $(a, c) \in \Sigma \times \Delta$. It is evident that the first property holds, so it remains to verify the second.

Define $\mathcal{Z} = \mathbb{C}^{\Sigma}$ and $\mathcal{W} = \mathbb{C}^{\Delta}$, and consider three registers Y, Z, and W corresponding to the spaces $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$, respectively. For the density operator $\rho \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{W})$ defined as

$$\rho = \sum_{(a,c) \in \Sigma \times \Delta} \lambda_{a,c} \Phi\big(x_{a,c} x_{a,c}^*\big) \otimes E_{a,a} \otimes E_{c,c}, \tag{8.19}$$

one has that the following two equalities hold:

$$\begin{aligned} \chi(\Phi(\tau)) &= \mathrm{D}\big(\rho[\mathsf{Y},\mathsf{Z},\mathsf{W}]\big\|\rho[\mathsf{Y}]\otimes\rho[\mathsf{Z},\mathsf{W}]\big), \\ \chi(\Phi(\eta)) &= \mathrm{D}\big(\rho[\mathsf{Y},\mathsf{Z}]\big\|\rho[\mathsf{Y}]\otimes\rho[\mathsf{Z}]\big). \end{aligned} \tag{8.20}$$

The inequality $\chi(\Phi(\eta)) \leq \chi(\Phi(\tau))$ follows from the monotonicity of the quantum relative entropy function under partial tracing (which represents a special case of Theorem 5.38). □

Theorem 8.10. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel, and let Σ be an alphabet having size $|\Sigma| = \dim(\mathcal{X})^2$. There exists an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ such that*

$$\chi(\Phi(\eta)) = \chi(\Phi). \tag{8.21}$$

One may assume, in addition, that $\mathrm{rank}(\eta(a)) \leq 1$ for each $a \in \Sigma$.

Proof. Consider an arbitrary ensemble of the form $\tau : \Gamma \to \mathrm{Pos}(\mathcal{X})$, for Γ being any alphabet, and let

$$\sigma = \sum_{a\in\Gamma} \tau(a) \tag{8.22}$$

denote the average state of the ensemble τ. Through Proposition 2.52, one finds that there must exist an alphabet Δ, a probability vector $p \in \mathcal{P}(\Delta)$, and a collection of ensembles $\{\tau_b : b \in \Delta\}$ taking the form $\tau_b : \Gamma \to \mathrm{Pos}(\mathcal{X})$, each satisfying the constraint

$$\sum_{a\in\Gamma} \tau_b(a) = \sigma \tag{8.23}$$

and possessing the property

$$\big|\{a \in \Gamma : \tau_b(a) \neq 0\}\big| \leq \dim(\mathcal{X})^2, \tag{8.24}$$

so that τ is given by the convex combination

$$\tau = \sum_{b\in\Delta} p(b)\,\tau_b. \tag{8.25}$$

By Proposition 5.51 it follows that

$$\chi(\Phi(\tau)) \leq \sum_{b\in\Delta} p(b)\chi(\Phi(\tau_b)), \tag{8.26}$$

and so there must exist at least one choice of a symbol $b \in \Delta$ for which $p(b) > 0$ and

$$\chi(\Phi(\tau)) \leq \chi(\Phi(\tau_b)). \tag{8.27}$$

Fix any such choice of $b \in \Delta$, and let

$$\Gamma_0 = \{a \in \Gamma \,:\, \tau_b(a) \neq 0\}. \tag{8.28}$$

For an arbitrarily chosen one-to-one mapping $f : \Gamma_0 \to \Sigma$, one obtains an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ such that

$$\chi(\Phi(\eta)) \geq \chi(\Phi(\tau)) \tag{8.29}$$

by setting $\eta(f(a)) = \tau_b(a)$ for every $a \in \Gamma_0$ and $\eta(c) = 0$ for $c \notin f(\Gamma_0)$.

Because the argument just presented holds for an arbitrary choice of an ensemble τ, it follows that

$$\chi(\Phi) = \sup_{\eta} \chi(\Phi(\eta)), \tag{8.30}$$

where the supremum is over all ensembles of the form $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$. As the set of all such ensembles is compact, there must exist an ensemble of the same form for which the equality (8.21) holds.

The additional restriction that $\mathrm{rank}(\eta(a)) \leq 1$ for each $a \in \Sigma$ may be assumed by first using Proposition 8.9 to replace a given ensemble τ by one satisfying the restriction $\mathrm{rank}(\tau(a)) \leq 1$ for each $a \in \Gamma$, and then proceeding with the argument above. This results in an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ with $\mathrm{rank}(\eta(a)) \leq 1$ for each $a \in \Sigma$, and such that (8.21) holds, which completes the proof. □

The entanglement-assisted Holevo capacity of a channel

Along similar lines to the entanglement-assisted classical capacity, which mirrors the definition of the classical capacity in the setting in which the sender and receiver initially share an entangled state of their choosing, one may define the entanglement-assisted Holevo capacity of a channel. The following definition is helpful when formalizing this notion.

Definition 8.11. Let Σ be an alphabet, let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{Y})$ be an ensemble, and let

$$\rho = \sum_{a \in \Sigma} \eta(a) \tag{8.31}$$

denote the average state of η. It is said that η is *constant with respect to* $\mathcal{Y}$ if and only if there exists a probability vector $p \in \mathcal{P}(\Sigma)$ such that

$$\operatorname{Tr}_{\mathcal{X}}(\eta(a)) = p(a) \operatorname{Tr}_{\mathcal{X}}(\rho) \tag{8.32}$$

for each $a \in \Sigma$.

A simple operational characterization of ensembles constant with respect to a given complex Euclidean space is provided by the following proposition. In essence, it states that this sort of ensemble is one obtained by applying a randomly chosen channel to just one subsystem of a fixed bipartite state.

Proposition 8.12. *Let Σ be an alphabet, let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, and let $\eta : \Sigma \to \operatorname{Pos}(\mathcal{X} \otimes \mathcal{Y})$ be an ensemble. The following three statements are equivalent:*

1. *The ensemble η is constant with respect to $\mathcal{Y}$.*
2. *There exists a complex Euclidean space $\mathcal{Z}$, a state $\sigma \in \mathrm{D}(\mathcal{Z} \otimes \mathcal{Y})$, a probability vector $p \in \mathcal{P}(\Sigma)$, and a collection of channels $\{\Phi_a : a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{Z}, \mathcal{X})$ such that*
$$\eta(a) = p(a)\big(\Phi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(\sigma) \tag{8.33}$$
for every $a \in \Sigma$.
3. *Statement 2 holds under the additional assumption that $\sigma = uu^*$ for some choice of a unit vector $u \in \mathcal{Z} \otimes \mathcal{Y}$.*

Proof. The fact that the second statement implies the first is immediate, and the third statement trivially implies the second. It therefore remains to prove that the first statement implies the third.

Assume that η is constant with respect to $\mathcal{Y}$, let ρ denote the average state of the ensemble η, as in Definition 8.11, and let

$$\xi = \operatorname{Tr}_{\mathcal{X}}(\rho). \tag{8.34}$$

Let $\mathcal{Z}$ be a complex Euclidean space having dimension $\dim(\mathcal{Z}) = \operatorname{rank}(\xi)$, and let $u \in \mathcal{Z} \otimes \mathcal{Y}$ be a unit vector that purifies ξ:

$$\operatorname{Tr}_{\mathcal{Z}}(uu^*) = \xi. \tag{8.35}$$

As η is constant with respect to $\mathcal{Y}$, it therefore holds that

$$p(a) \operatorname{Tr}_{\mathcal{Z}}(uu^*) = \operatorname{Tr}_{\mathcal{X}}(\eta(a)) \tag{8.36}$$

for each $a \in \Sigma$. By Proposition 2.29, one concludes that there must exist a channel $\Phi_a \in \mathrm{C}(\mathcal{Z}, \mathcal{X})$ such that

$$\eta(a) = p(a)\big(\Psi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}\big)\big(uu^*\big) \tag{8.37}$$

for each $a \in \Sigma$. Setting $\sigma = uu^*$ completes the proof. □

Definition 8.13. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. The *entanglement-assisted Holevo capacity* of Φ is defined as

$$\chi_{\mathrm{E}}(\Phi) = \sup_{\eta} \chi\big(\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\eta)\big), \tag{8.38}$$

where the supremum is taken over all choices of an alphabet Σ, a complex Euclidean space $\mathcal{W}$, and an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{W})$ that is constant with respect to $\mathcal{W}$.

The relationship between the entanglement-assisted classical capacity and the entanglement-assisted Holevo capacity is discussed in Section 8.1.3 below. In this context, the fixed bipartite state whose existence is implied by Proposition 8.12, for a given ensemble that is constant with respect to $\mathcal{W}$, is representative of a (possibly entangled) state shared between a sender and receiver.

The coherent information

The final quantity, associated with a given channel, that is to be defined in the present subsection is the *coherent information*.

Definition 8.14. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel and let $\sigma \in \mathrm{D}(\mathcal{X})$ be a state, for $\mathcal{X}$ and $\mathcal{Y}$ being complex Euclidean spaces. The *coherent information* of σ through Φ is defined as

$$\mathrm{I}_{\mathrm{C}}(\sigma\,;\,\Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}\Big(\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)\Big(\mathrm{vec}\big(\sqrt{\sigma}\big)\,\mathrm{vec}\big(\sqrt{\sigma}\big)^*\Big)\Big). \tag{8.39}$$

The *maximum coherent information* of Φ is defined as

$$\mathrm{I}_{\mathrm{C}}(\Phi) = \max_{\sigma \in \mathrm{D}(\mathcal{X})} \mathrm{I}_{\mathrm{C}}(\sigma;\Phi). \tag{8.40}$$

In general terms, the coherent information of a state σ through a channel Φ quantifies the correlations that exist after Φ is applied to a purification

of σ. The definition implicitly takes this purification to be $\operatorname{vec}(\sqrt{\sigma})$ for the sake of simplicity and concreteness, but any other purification would result in the same quantity.

Consider the state

$$\rho = \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)\Big(\operatorname{vec}\big(\sqrt{\sigma}\big)\operatorname{vec}\big(\sqrt{\sigma}\big)^*\Big) \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{X}) \tag{8.41}$$

of a pair of registers (Y, X), corresponding to the spaces $\mathcal{Y}$ and $\mathcal{X}$, obtained as suggested above. One has that the coherent information $\mathrm{I}_{\mathrm{c}}(\sigma\,;\,\Phi)$ of σ through Φ is equal to $\mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{Y},\mathsf{X})$. The mutual information between Y and X is therefore given by

$$\mathrm{H}(\mathsf{Y} : \mathsf{X}) = \mathrm{I}_{\mathrm{c}}(\sigma\,;\,\Phi) + \mathrm{H}(\sigma). \tag{8.42}$$

While it is not immediately clear that the coherent information is relevant to the notion of channel capacity, it will be proved later in the chapter that this quantity is fundamentally important with respect to both the entanglement-assisted classical capacity and the quantum capacity (to be defined later in Section 8.2).

The following proposition establishes an intuitive fact, which is that feeding the output of a channel into a second channel cannot increase the first channel's coherent information relative to a given state.

Proposition 8.15. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{Y},\mathcal{Z})$ be channels, and let $\sigma \in \mathrm{D}(\mathcal{X})$ be a state. It holds that*

$$\mathrm{I}_{\mathrm{c}}(\sigma\,;\,\Psi\Phi) \leq \mathrm{I}_{\mathrm{c}}(\sigma\,;\,\Phi). \tag{8.43}$$

Proof. Choose complex Euclidean spaces $\mathcal{W}$ and $\mathcal{V}$, along with isometries $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y}\otimes\mathcal{W})$ and $B \in \mathrm{U}(\mathcal{Y}, \mathcal{Z}\otimes\mathcal{V})$, so that Stinespring representations of Φ and Ψ are obtained:

$$\Phi(X) = \operatorname{Tr}_{\mathcal{W}}(AXA^*) \quad \text{and} \quad \Psi(Y) = \operatorname{Tr}_{\mathcal{V}}(BYB^*) \tag{8.44}$$

for all $X \in \mathrm{L}(\mathcal{X})$ and $Y \in \mathrm{L}(\mathcal{Y})$. Define a unit vector $u \in \mathcal{Z}\otimes\mathcal{V}\otimes\mathcal{W}\otimes\mathcal{X}$ as

$$u = (B \otimes \mathbb{1}_{\mathcal{W}} \otimes \mathbb{1}_{\mathcal{X}})(A \otimes \mathbb{1}_{\mathcal{X}})\operatorname{vec}\big(\sqrt{\sigma}\big). \tag{8.45}$$

Now, consider four registers $\mathsf{Z}, \mathsf{V}, \mathsf{W}$, and X, corresponding to the spaces $\mathcal{Z}$, $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{X}$, respectively. Assuming the compound register $(\mathsf{Z},\mathsf{V},\mathsf{W},\mathsf{X})$ is in the pure state uu^*, one has the following expressions:

$$\begin{aligned} \mathrm{I}_{\mathrm{c}}(\sigma\,;\,\Phi) &= \mathrm{H}(\mathsf{Z},\mathsf{V}) - \mathrm{H}(\mathsf{Z},\mathsf{V},\mathsf{X}), \\ \mathrm{I}_{\mathrm{c}}(\sigma\,;\,\Psi\Phi) &= \mathrm{H}(\mathsf{Z}) - \mathrm{H}(\mathsf{Z},\mathsf{X}). \end{aligned} \tag{8.46}$$

The proposition follows from the strong subadditivity of the von Neumann entropy (Theorem 5.39). □

It is convenient to refer to the notion of *complementary channels* in some of the proofs to be found in the present chapter that involve the coherent information. This notion is defined as follows.

Definition 8.16. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be channels, for some choice of complex Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. It is said that Φ and Ψ are *complementary* if and only if there exists an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ for which it holds that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \quad \text{and} \quad \Psi(X) = \mathrm{Tr}_{\mathcal{Y}}(AXA^*) \tag{8.47}$$

for every $X \in \mathrm{L}(\mathcal{X})$.

It is immediate from Corollary 2.27 that, for every channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, there must exist a complex Euclidean space $\mathcal{Z}$ and a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ that is complementary to Φ; such a channel Ψ is obtained from any choice of a Stinespring representation of Φ.

Proposition 8.17. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let $\sigma \in \mathrm{D}(\mathcal{X})$ be a density operator. If $\mathcal{Z}$ is a complex Euclidean space and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ is a channel that is complementary to Φ, then*

$$\mathrm{I}_{\mathrm{C}}(\sigma\,;\Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)). \tag{8.48}$$

Proof. Under the assumption that $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ is complementary to Φ, there must exist an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ such that the equations (8.47) hold for every $X \in \mathrm{L}(\mathcal{X})$. Let X, Y, and Z be registers corresponding to the spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, define a unit vector $u \in \mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{X}$ as

$$u = \bigl(A \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\bigr)\,\mathrm{vec}\bigl(\sqrt{\sigma}\bigr), \tag{8.49}$$

and consider the compound register $(\mathsf{Y}, \mathsf{Z}, \mathsf{X})$. With respect to the pure state uu^* of this compound register, it holds that $\mathrm{H}(\mathsf{Z}) = \mathrm{H}(\mathsf{Y}, \mathsf{X})$, and therefore

$$\mathrm{H}\Bigl(\bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\bigr)\Bigl(\mathrm{vec}\bigl(\sqrt{\sigma}\bigr)\,\mathrm{vec}\bigl(\sqrt{\sigma}\bigr)^*\Bigr)\Bigr) = \mathrm{H}(\Psi(\sigma)), \tag{8.50}$$

from which the proposition follows. □

8.1.2 The Holevo–Schumacher–Westmoreland theorem

This section states and proves the *Holevo–Schumacher–Westmoreland theorem*, which establishes that the classical capacity of a quantum channel is lower-bounded by its Holevo capacity, and that through a regularization of the Holevo capacity one obtains a characterization of the classical capacity.

The notion of a *classical-to-quantum product state channel code*, along with a few mathematical results that are useful for analyzing these codes, will be introduced prior to the statement and proof of the Holevo–Schumacher–Westmoreland theorem.

Classical-to-quantum product state channel codes

When studying the classical capacity of quantum channels, it is instructive to consider a related but somewhat more basic task of encoding classical information using fixed sets of quantum states. When this task is connected with the notion of the classical capacity of a given channel, a link must be made between that particular channel and the states that are used to encode classical information—but it is reasonable to begin by examining the task of encoding classical information into quantum states in isolation.

Throughout the discussion that follows, $\Gamma = \{0,1\}$ will denote the binary alphabet and

$$\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.51}$$

will denote a fixed collection of states, for $\mathcal{X}$ being a complex Euclidean space and Σ being an alphabet.[1] The situation to be considered is that binary strings, representing classical information, are to be encoded into tensor products of quantum states drawn from the collection (8.51) in such a way that each binary string can be recovered from its encoding with high probability.

In more precise terms, it is to be assumed that positive integers n and m have been selected, and that every binary string $b_1 \cdots b_m \in \Gamma^m$ of length m is to be *encoded* by a product state having the form

$$\sigma_{a_1} \otimes \cdots \otimes \sigma_{a_n} \in \mathrm{D}\big(\mathcal{X}^{\otimes n}\big), \tag{8.52}$$

[1] The entire discussion could be generalized to allow for arbitrary alphabets Γ in place of the binary alphabet. As there is little gain in doing this from the perspective of this book, the assumption that $\Gamma = \{0,1\}$ is made in the interest of simplicity.

for some choice of a string $a_1 \cdots a_n \in \Sigma^n$. That is, a function $f : \Gamma^m \to \Sigma^n$ is to be selected, and each string $b_1 \cdots b_m \in \Gamma^m$ is to be encoded by the state (8.52) for $a_1 \cdots a_n = f(b_1 \cdots b_m)$. When discussing this sort of code, it is convenient to make use of the shorthand notation

$$\sigma_{a_1 \cdots a_n} = \sigma_{a_1} \otimes \cdots \otimes \sigma_{a_n} \tag{8.53}$$

for each string $a_1 \cdots a_n \in \Sigma^n$, and with respect to this notation one has that

$$\sigma_{f(b_1 \cdots b_m)} \in \mathrm{D}\big(\mathcal{X}^{\otimes n}\big) \tag{8.54}$$

denotes the state that encodes the string $b_1 \cdots b_m \in \Gamma^m$.

From the encoding of a given binary string, one may hope to recover (or *decode*) this string by means of a measurement. Such a measurement takes the form $\mu : \Gamma^m \to \mathrm{Pos}\big(\mathcal{X}^{\otimes n}\big)$, and succeeds in successfully recovering a particular string $b_1 \cdots b_m$ from its encoding with probability

$$\big\langle \mu(b_1 \cdots b_m), \sigma_{f(b_1 \cdots b_m)} \big\rangle. \tag{8.55}$$

As a general guideline, one is typically interested in coding schemes for which the probability of a successful decoding is close to 1 and the ratio m/n, which represents the rate at which classical information is effectively transmitted, is as large as possible. The following definition summarizes these notions.

Definition 8.18. Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let

$$\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.56}$$

be a collection of states, let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let n and m be positive integers. A *classical-to-quantum product-state channel code* for the collection of states (8.56) is a pair (f, μ) consisting of a function and a measurement of the forms

$$f : \Gamma^m \to \Sigma^n \quad \text{and} \quad \mu : \Gamma^m \to \mathrm{Pos}\big(\mathcal{X}^{\otimes n}\big). \tag{8.57}$$

The *rate* of such a code is equal to the ratio m/n, and the code is said to have *error bounded by* δ if it holds that

$$\big\langle \mu(b_1 \cdots b_m), \sigma_{f(b_1 \cdots b_m)} \big\rangle > 1 - \delta \tag{8.58}$$

for every string $b_1 \cdots b_m \in \Gamma^m$.

Remark 8.19. The term *channel code* is used in this definition to distinguish this type of code from a *source code*, as discussed in Chapter 5. The two notions are, in some sense, complementary. A channel code represents the situation in which information is encoded into a state possessing a degree of noise or randomness, while a source code represents the situation in which information produced by a noisy or random source is encoded into any chosen state.

It is evident that some choices of sets $\{\sigma_a : a \in \Sigma\}$ are better suited to the construction of classical-to-quantum product state channel codes than others, assuming one wishes to maximize the rate and minimize the error probability for such a code. For the most part, the analysis that follows will be focused on the situation in which a set of states has been fixed, and one is interested in understanding the capabilities of this particular set, with respect to classical-to-quantum product state channel codes.

Typicality for ensembles of states

The notion of *typicality* is central to the proofs of multiple theorems to be presented in the current chapter, including a fundamental theorem on the existence of classical-to-quantum product-state channel codes possessing certain rates and error bounds.

A standard definition of typicality was introduced in Section 5.3.1—but it is an extension of this definition to ensembles of states that will be used in the context of channel coding. The following definition is a starting point for a discussion of this concept, providing a notion of typicality for joint probability distributions.

Definition 8.20. Let Σ and Γ be alphabets, let $p \in \mathcal{P}(\Sigma \times \Gamma)$ be a probability vector, and let $q \in \mathcal{P}(\Sigma)$ be the marginal probability vector defined as

$$q(a) = \sum_{b \in \Gamma} p(a, b) \tag{8.59}$$

for each $a \in \Sigma$. For each choice of a positive real number $\varepsilon > 0$, a positive integer n, and a string $a_1 \cdots a_n \in \Sigma^n$ satisfying $q(a_1) \cdots q(a_n) > 0$, a string $b_1 \cdots b_n \in \Gamma^n$ is said to be *ε-typical conditioned on* $a_1 \cdots a_n \in \Sigma^n$ if and only if

$$2^{-n(\mathrm{H}(p)-\mathrm{H}(q)+\varepsilon)} < \frac{p(a_1, b_1) \cdots p(a_n, b_n)}{q(a_1) \cdots q(a_n)} < 2^{-n(\mathrm{H}(p)-\mathrm{H}(q)-\varepsilon)}. \tag{8.60}$$

One writes $K_{a_1 \cdots a_n, \varepsilon}(p)$ to denote the set of all such strings $b_1 \cdots b_n \in \Gamma^n$.

When a joint probability vector $p \in \mathcal{P}(\Sigma \times \Gamma)$ is fixed, or can safely be taken as being implicit, the notation $K_{a_1 \cdots a_n, \varepsilon}$ may be used in place of $K_{a_1 \cdots a_n, \varepsilon}(p)$. It is also convenient to define $K_{a_1 \cdots a_n, \varepsilon}(p) = \varnothing$ for any string $a_1 \cdots a_n \in \Sigma^n$ for which $q(a_1) \cdots q(a_n) = 0$.

Intuitively speaking, if one were to select strings $a_1 \cdots a_n \in \Sigma^n$ and $b_1 \cdots b_n \in \Gamma^n$ by choosing the pairs $(a_1, b_1), \ldots, (a_n, b_n)$ independently at random according to a probability vector $p \in \mathcal{P}(\Sigma \times \Gamma)$, then it is reasonable to expect that $b_1 \cdots b_n$ will be contained in $K_{a_1 \cdots a_n, \varepsilon}(p)$, with the probability of this event becoming increasingly likely as n becomes large. This fact is established by the following proposition, based on the weak law of large numbers (Theorem 1.19)—the methodology is essentially the same as the analogous fact (Proposition 5.45) that was proved in regard to the standard definition of typicality discussed in Section 5.3.1.

Proposition 8.21. *Let Σ and Γ be alphabets, let $p \in \mathcal{P}(\Sigma \times \Gamma)$ be a probability vector, and let $q \in \mathcal{P}(\Sigma)$ be the marginal probability vector defined as*

$$q(a) = \sum_{b \in \Gamma} p(a, b) \tag{8.61}$$

for each $a \in \Sigma$. For every $\varepsilon > 0$ it holds that

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in \Sigma^n} \sum_{b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}} p(a_1, b_1) \cdots p(a_n, b_n) = 1. \tag{8.62}$$

Proof. Let n be a positive integer and let $X_1, \ldots, X_n$ be independent and identically distributed random variables, defined as

$$X_k(a, b) = -\log(p(a, b)) + \log(q(a)) \tag{8.63}$$

for each pair $(a, b) \in \Sigma \times \Gamma$, and distributed with respect to p. One has that the expected value $\alpha = \mathrm{E}(X_k)$ of each of these random variables is given by $\alpha = \mathrm{H}(p) - \mathrm{H}(q)$, and furthermore

$$\begin{aligned} &\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - \alpha \right| < \varepsilon \right) \\ &\qquad = \sum_{a_1 \cdots a_n \in \Sigma^n} \sum_{b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}} p(a_1, b_1) \cdots p(a_n, b_n). \end{aligned} \tag{8.64}$$

The conclusion of the proposition therefore follows from the weak law of large numbers (Theorem 1.19). □

The next proposition places an upper bound on the expected size of the set $K_{a_1\cdots a_n,\varepsilon}$. It is analogous to Proposition 5.46 for the standard definition of typicality.

Proposition 8.22. *Let Σ and Γ be alphabets, let $p \in \mathcal{P}(\Sigma \times \Gamma)$ be a probability vector, and let $q \in \mathcal{P}(\Sigma)$ be the marginal probability vector defined as*

$$q(a) = \sum_{b\in\Gamma} p(a,b) \tag{8.65}$$

for each $a \in \Sigma$. For every positive integer n and every positive real number $\varepsilon > 0$, it holds that

$$\sum_{a_1\cdots a_n\in\Sigma^n} q(a_1)\cdots q(a_n)\left|K_{a_1\cdots a_n,\varepsilon}\right| < 2^{n(\mathrm{H}(p)-\mathrm{H}(q)+\varepsilon)}. \tag{8.66}$$

Proof. For each string $a_1\cdots a_n \in \Sigma^n$ satisfying $q(a_1)\cdots q(a_n) > 0$ and each string $b_1\cdots b_n \in K_{a_1\cdots a_n,\varepsilon}$, one has

$$2^{-n(\mathrm{H}(p)-\mathrm{H}(q)+\varepsilon)} < \frac{p(a_1,b_1)\cdots p(a_n,b_n)}{q(a_1)\cdots q(a_n)}, \tag{8.67}$$

and therefore

$$\begin{aligned} &2^{-n(\mathrm{H}(p)-\mathrm{H}(q)+\varepsilon)} \sum_{a_1\cdots a_n\in\Sigma^n} q(a_1)\cdots q(a_n)|K_{a_1\cdots a_n,\varepsilon}| \\ &\quad< \sum_{a_1\cdots a_n\in\Sigma^n} \sum_{b_1\cdots b_n\in K_{a_1\cdots a_n,\varepsilon}} p(a_1,b_1)\cdots p(a_n,b_n) \leq 1, \end{aligned} \tag{8.68}$$

from which the proposition follows. □

The notion of typicality for joint probability distributions established by Definition 8.20 may be extended to ensembles of quantum states in a fairly straightforward fashion, using spectral decompositions of the states in an ensemble.

Definition 8.23. Let $\mathcal{X}$ be a complex Euclidean space, let Σ be an alphabet, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states, and let Γ be an alphabet such that $|\Gamma| = \dim(\mathcal{X})$. By the spectral theorem (as stated by Corollary 1.4), one may write

$$\eta(a) = \sum_{b\in\Gamma} p(a,b)\, u_{a,b} u_{a,b}^* \tag{8.69}$$

for $p \in \mathcal{P}(\Sigma \times \Gamma)$ being a probability vector and $\{u_{a,b} : b \in \Gamma\}$ being an orthonormal basis of $\mathcal{X}$ for each $a \in \Sigma$. With respect to the ensemble η, and for each positive real number $\varepsilon > 0$, each positive integer n, and each string $a_1 \cdots a_n \in \Sigma^n$, the *projection onto the ε-typical subspace of $\mathcal{X}^{\otimes n}$ conditioned on $a_1 \cdots a_n$* is defined as

$$\Lambda_{a_1 \cdots a_n, \varepsilon} = \sum_{b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}} u_{a_1, b_1} u_{a_1, b_1}^* \otimes \cdots \otimes u_{a_n, b_n} u_{a_n, b_n}^*. \tag{8.70}$$

Remark 8.24. For a fixed choice of a string $a_1 \cdots a_n \in \Sigma^n$, one has that the inclusion of each string $b_1 \cdots b_n$ in $K_{a_1 \cdots a_n, \varepsilon}$ is determined by the multiset of values $\{p(a_1, b_1), \ldots, p(a_n, b_n)\}$ alone. Thus, the same is true regarding the inclusion of each rank-one projection in the summation (8.70). It follows that the projection $\Lambda_{a_1 \cdots a_n, \varepsilon}$ specified by Definition 8.23 is uniquely defined by the ensemble η, and is independent of the particular choices of the spectral decompositions (8.69).

Facts analogous to the previous two propositions, holding for ensembles rather than joint probability distributions, follow directly.

Proposition 8.25. *Let Σ be an alphabet, let $\mathcal{X}$ be a complex Euclidean space, and let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states. For every $\varepsilon > 0$, it holds that*

$$\lim_{n \to \infty} \sum_{a_1 \cdots a_n \in \Sigma^n} \left\langle \Lambda_{a_1 \cdots a_n, \varepsilon}, \eta(a_1) \otimes \cdots \otimes \eta(a_n) \right\rangle = 1, \tag{8.71}$$

where, for each positive integer n, and each string $a_1 \cdots a_n \in \Sigma^n$, $\Lambda_{a_1 \cdots a_n, \varepsilon}$ is the projection onto the ε-typical subspace of $\mathcal{X}^{\otimes n}$ conditioned on $a_1 \cdots a_n$, with respect to the ensemble η. Moreover, one has

$$\sum_{a_1 \cdots a_n \in \Sigma^n} \mathrm{Tr}(\eta(a_1)) \cdots \mathrm{Tr}(\eta(a_n)) \ \mathrm{Tr}\left(\Lambda_{a_1 \cdots a_n, \varepsilon}\right) < 2^{n(\beta + \varepsilon)} \tag{8.72}$$

for

$$\beta = \sum_{\substack{a \in \Sigma \\ \eta(a) \neq 0}} \mathrm{Tr}(\eta(a)) \ \mathrm{H}\left(\frac{\eta(a)}{\mathrm{Tr}(\eta(a))}\right). \tag{8.73}$$

Proof. Assume that

$$\eta(a) = \sum_{b \in \Gamma} p(a, b) u_{a,b} u_{a,b}^* \tag{8.74}$$

is a spectral decomposition of $\eta(a)$ for each $a \in \Sigma$, and define $q \in \mathcal{P}(\Sigma)$ as $q(a) = \sum_{b \in \Gamma} p(a,b)$ for each $a \in \Sigma$ (which is equivalent to $q(a) = \mathrm{Tr}(\eta(a))$ for each $a \in \Sigma$). For each positive integer n, each positive real number $\varepsilon > 0$, and each string $a_1 \cdots a_n \in \Sigma^n$, one has

$$\begin{aligned} &\bigl\langle \Lambda_{a_1 \cdots a_n, \varepsilon}, \eta(a_1) \otimes \cdots \otimes \eta(a_n) \bigr\rangle \\ &\qquad = \sum_{b_1 \cdots b_n \in K_{a_1 \cdots a_n, \varepsilon}} p(a_1, b_1) \cdots p(a_n, b_n), \end{aligned} \tag{8.75}$$

and moreover

$$\beta = \mathrm{H}(p) - \mathrm{H}(q) \quad \text{and} \quad \mathrm{Tr}\bigl(\Lambda_{a_1 \cdots a_n, \varepsilon}\bigr) = \bigl|K_{a_1 \cdots a_n, \varepsilon}\bigr|. \tag{8.76}$$

The proposition therefore follows from Propositions 8.21 and 8.22. □

A useful operator inequality

When analyzing the performance of classical-to-quantum product state channel codes, it is helpful to make use of the operator inequality to be stated as Lemma 8.28 below. The proof of this inequality will make use of the following fact regarding square roots of positive semidefinite operators.

Lemma 8.26 (Operator monotonicity of the square root)**.** *Let $\mathcal{X}$ be a complex Euclidean space and let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators. It holds that*

$$\sqrt{P} \leq \sqrt{P + Q}. \tag{8.77}$$

Proof. The block operator

$$\begin{pmatrix} P & \sqrt{P} \\ \sqrt{P} & \mathbb{1} \end{pmatrix} + \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} P + Q & \sqrt{P} \\ \sqrt{P} & \mathbb{1} \end{pmatrix} \tag{8.78}$$

is positive semidefinite. As $[P + Q, \mathbb{1}] = 0$ and $\sqrt{P}$ is Hermitian, it follows by Lemma 5.31 that

$$\sqrt{P} \leq \sqrt{P + Q}\sqrt{\mathbb{1}} = \sqrt{P + Q}, \tag{8.79}$$

as required. □

Remark 8.27. It is not difficult to prove Lemma 8.26 directly, without relying on Lemma 5.31, by using spectral properties of operators that were also employed in the proof of that lemma.

Lemma 8.28 (Hayashi–Nagaoka). *Let $P, Q \in \mathrm{Pos}(\mathcal{X})$ be positive semidefinite operators, for $\mathcal{X}$ being a complex Euclidean space, and assume $P \leq \mathbb{1}$. It holds that*

$$\mathbb{1} - \sqrt{(P+Q)^+}\, P \sqrt{(P+Q)^+} \leq 2(\mathbb{1} - P) + 4Q. \tag{8.80}$$

Proof. For every choice of operators $A, B \in \mathrm{L}(\mathcal{X})$, one has

$$0 \leq (A-B)(A-B)^* = AA^* + BB^* - (AB^* + BA^*), \tag{8.81}$$

and therefore $AB^* + BA^* \leq AA^* + BB^*$. Setting

$$A = X\sqrt{Q} \quad \text{and} \quad B = (\mathbb{1} - X)\sqrt{Q}, \tag{8.82}$$

for a given operator $X \in \mathrm{L}(\mathcal{X})$, yields

$$XQ(\mathbb{1}-X)^* + (\mathbb{1}-X)QX^* \leq XQX^* + (\mathbb{1}-X)Q(\mathbb{1}-X)^*, \tag{8.83}$$

and therefore

$$\begin{aligned} Q &= XQX^* + XQ(\mathbb{1}-X)^* + (\mathbb{1}-X)QX^* + (\mathbb{1}-X)Q(\mathbb{1}-X)^* \\ &\leq 2XQX^* + 2(\mathbb{1}-X)Q(\mathbb{1}-X)^*. \end{aligned} \tag{8.84}$$

For the specific choice $X = \sqrt{P+Q}$, one obtains

$$Q \leq 2\sqrt{P+Q}\, Q \sqrt{P+Q} + 2\Big(\mathbb{1} - \sqrt{P+Q}\Big)\, Q \,\Big(\mathbb{1} - \sqrt{P+Q}\Big), \tag{8.85}$$

and from the observation that $Q \leq P + Q$ it follows that

$$\begin{aligned} Q &\leq 2\sqrt{P+Q}\, Q \sqrt{P+Q} \\ &\qquad + 2\Big(\mathbb{1} - \sqrt{P+Q}\Big)\, (P+Q)\, \Big(\mathbb{1} - \sqrt{P+Q}\Big) \\ &= \sqrt{P+Q}\Big(2\mathbb{1} + 4Q - 4\sqrt{P+Q} + 2P\Big)\sqrt{P+Q}. \end{aligned} \tag{8.86}$$

Using the fact that $P \leq \mathbb{1}$ together with Lemma 8.26, one has

$$P \leq \sqrt{P} \leq \sqrt{P+Q}, \tag{8.87}$$

and therefore

$$Q \leq \sqrt{P+Q}\Big(2\mathbb{1} - 2P + 4Q\Big)\sqrt{P+Q}. \tag{8.88}$$

Conjugating both sides of this inequality by the Moore–Penrose pseudo-inverse of $\sqrt{P+Q}$ yields

$$\sqrt{(P+Q)^+}\, Q \sqrt{(P+Q)^+} \leq 2\Pi_{\mathrm{im}(P+Q)} - 2P + 4Q. \tag{8.89}$$

It follows that

$$\begin{aligned} \mathbb{1} - \sqrt{(P+Q)^+}\, P \sqrt{(P+Q)^+} & \\ &= \mathbb{1} - \Pi_{\mathrm{im}(P+Q)} + \sqrt{(P+Q)^+}\, Q \sqrt{(P+Q)^+} \\ &\leq \mathbb{1} + \Pi_{\mathrm{im}(P+Q)} - 2P + 4Q \\ &\leq 2(\mathbb{1} - P) + 4Q, \end{aligned} \tag{8.90}$$

as required. □

An existence proof for classical-to-quantum product state channel codes

Returning to the discussion of classical-to-quantum product-state channel codes, assume as before that an alphabet Σ, a complex Euclidean space $\mathcal{X}$, and a set of states

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.91}$$

has been fixed, and let $\Gamma = \{0,1\}$ denote the binary alphabet. It is natural to ask, for any choice of a positive real number $\delta > 0$ and positive integers m and n, whether or not there exists a classical-to-quantum product-state channel code (f, μ) taking the form

$$f : \Gamma^m \to \Sigma^n \quad \text{and} \quad \mu : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes m}) \tag{8.92}$$

and having error bounded by δ.

In general, one may expect that making such a determination is not tractable from a computational point of view. It is possible, however, to prove the existence of reasonably good classical-to-quantum product-state channel codes through the probabilistic method: for suitable choices of n, m, and δ, a *random* choice of a function $f : \Gamma^m \to \Sigma^n$ and a well-chosen measurement $\mu : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes n})$ are considered, and a coding scheme having error bounded by δ is obtained with a nonzero probability. The next theorem gives a precise statement regarding the parameters n, m, and δ through which this methodology proves the existence of quantum-to-classical product-state channels codes.

Theorem 8.29. *Let Σ be an alphabet, let $\mathcal{X}$ be a complex Euclidean space, let*

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.93}$$

be a collection of states, and let $\Gamma = \{0,1\}$ denote the binary alphabet. Also let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be the ensemble defined as

$$\eta(a) = p(a)\sigma_a \tag{8.94}$$

for each $a \in \Sigma$, let α be a positive real number satisfying $\alpha < \chi(\eta)$, and let $\delta > 0$ be a positive real number. For all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, there exists a function $f : \Gamma^m \to \Sigma^n$ and a measurement $\mu : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes n})$ such that

$$\bigl\langle \mu(b_1 \cdots b_m), \sigma_{f(b_1 \cdots b_m)} \bigr\rangle > 1 - \delta \tag{8.95}$$

for every $b_1 \cdots b_m \in \Gamma^m$.

Proof. It will first be assumed that n and m are arbitrary positive integers. As suggested previously, the proof makes use of the probabilistic method: a random function $g : \Gamma^{m+1} \to \Sigma^n$ is chosen from a particular probability distribution, a decoding measurement μ is defined for each possible choice of g, and the expected probability of a decoding error for the pair (g, μ) is analyzed. As is to be explained later in the proof, this analysis implies the existence of a channel coding scheme (f, μ), where $f : \Gamma^m \to \Sigma^n$ is derived from g, satisfying the requirements theorem for all but finitely many n and for $m \leq \alpha n$.

The particular distribution from which g is to be chosen is one in which each individual output symbol of g is selected independently according to the probability vector p. Equivalently, for a random selection of g according to the distribution being described, one has that

$$\Pr\bigl(g(b_1 \cdots b_{m+1}) = a_1 \cdots a_n\bigr) = p(a_1) \cdots p(a_n) \tag{8.96}$$

for every choice of $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$ and $a_1 \cdots a_n \in \Sigma^n$, and moreover the outputs of a randomly chosen g on distinct choices of the input string $b_1 \cdots b_{m+1}$ are uncorrelated.

The specification of the decoding measurement μ that is to be associated with a given g is not chosen randomly—a unique measurement is defined

for each g in a way that is dependent upon the ensemble η. First, let $\varepsilon > 0$ be a sufficiently small positive real number such that the inequality

$$\alpha < \chi(\eta) - 3\varepsilon \tag{8.97}$$

holds. For each string $a_1 \cdots a_n \in \Sigma^n$, let $\Lambda_{a_1 \cdots a_n}$ denote the projection onto the ε-typical subspace of $\mathcal{X}^{\otimes n}$ conditioned on $a_1 \cdots a_n$, with respect to the ensemble η, and let Π_n be the projection onto the ε-typical subspace of $\mathcal{X}^{\otimes n}$ with respect to the average state

$$\sigma = \sum_{a \in \Sigma} p(a)\sigma_a \tag{8.98}$$

of the ensemble η. (As ε has been fixed, the dependence of $\Lambda_{a_1 \cdots a_n}$ and Π_n on ε is not written explicitly, allowing for slightly less cluttered equations.) Next, for a given choice of $g : \Gamma^{m+1} \to \Sigma^n$, define

$$Q = \sum_{b_1 \cdots b_{m+1} \in \Gamma^{m+1}} \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n \,, \tag{8.99}$$

and, for each binary string $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$, define

$$Q_{b_1 \cdots b_{m+1}} = \sqrt{Q^+}\, \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n \sqrt{Q^+}. \tag{8.100}$$

One has that each operator $Q_{b_1 \cdots b_{m+1}}$ is positive semidefinite, and moreover

$$\sum_{b_1 \cdots b_{m+1} \in \Gamma^{m+1}} Q_{b_1 \cdots b_{m+1}} = \Pi_{\mathrm{im}(Q)}. \tag{8.101}$$

Finally, the measurement $\mu : \Gamma^{m+1} \to \mathrm{Pos}\big(\mathcal{X}^{\otimes n}\big)$ to be associated with g is defined as

$$\mu(b_1 \cdots b_{m+1}) = Q_{b_1 \cdots b_{m+1}} + \frac{1}{2^{m+1}} \big(\mathbb{1} - \Pi_{\mathrm{im}(Q)}\big) \tag{8.102}$$

for each $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$.

For each choice of g, the probability that the measurement μ associated with g errs in recovering a string $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$ from its encoding is equal to

$$\big\langle \mathbb{1} - \mu(b_1 \cdots b_{m+1}), \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle. \tag{8.103}$$

The next phase of the proof establishes a lower-bound on the average error probability

$$\frac{1}{2^{m+1}} \sum_{b_1 \cdots b_{m+1} \in \Gamma^{m+1}} \left\langle \mathbb{1} - \mu(b_1 \cdots b_{m+1}), \sigma_{g(b_1 \cdots b_{m+1})} \right\rangle, \tag{8.104}$$

for a uniformly chosen string $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$. To bound this average probability of error, one may first observe that Lemma 8.28 implies that

$$\begin{aligned} &\mathbb{1} - Q_{b_1 \cdots b_{m+1}} \\ &\quad \leq 2\left(\mathbb{1} - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n\right) + 4\left(Q - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n\right) \end{aligned} \tag{8.105}$$

for each $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$. For a fixed choice of g, the probability of an error in recovering a given string $b_1 \cdots b_{m+1}$ is therefore upper-bounded by

$$\begin{aligned} &2\left\langle \mathbb{1} - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n, \sigma_{g(b_1 \cdots b_{m+1})} \right\rangle \\ &\quad + 4\left\langle Q - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n, \sigma_{g(b_1 \cdots b_{m+1})} \right\rangle. \end{aligned} \tag{8.106}$$

The expected value of this expression will be shown to be small, under the additional assumption that $m \leq \alpha n$, when $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$ is chosen uniformly and g is chosen according to the distribution described above.

The first term in the expression (8.106) will be considered first. To prove an upper bound on the expected value of this quantity, it is convenient to make use of the operator identity

$$ABA = AB + BA - B + (\mathbb{1} - A)B(\mathbb{1} - A). \tag{8.107}$$

In particular, for any choice of a string $a_1 \cdots a_n \in \Sigma^n$, this identity implies

$$\begin{aligned} &\left\langle \Pi_n \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \right\rangle \\ &\quad = \left\langle \Pi_n \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \right\rangle + \left\langle \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \right\rangle - \left\langle \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \right\rangle \\ &\qquad + \left\langle (\mathbb{1} - \Pi_n) \Lambda_{a_1 \cdots a_n} (\mathbb{1} - \Pi_n), \sigma_{a_1 \cdots a_n} \right\rangle \\ &\quad \geq \left\langle \Pi_n \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \right\rangle + \left\langle \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \right\rangle - \left\langle \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \right\rangle. \end{aligned} \tag{8.108}$$

As $\Lambda_{a_1 \cdots a_n}$ is a projection operator and commutes with $\sigma_{a_1 \cdots a_n}$, it follows that

$$\begin{aligned} &\left\langle \Pi_n \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \right\rangle \\ &\quad \geq \left\langle 2\Pi_n - \mathbb{1}, \Lambda_{a_1 \cdots a_n} \sigma_{a_1 \cdots a_n} \right\rangle \\ &\quad = \left\langle 2\Pi_n - \mathbb{1}, \sigma_{a_1 \cdots a_n} \right\rangle + \left\langle \mathbb{1} - 2\Pi_n, (\mathbb{1} - \Lambda_{a_1 \cdots a_n}) \sigma_{a_1 \cdots a_n} \right\rangle \\ &\quad \geq \left\langle 2\Pi_n - \mathbb{1}, \sigma_{a_1 \cdots a_n} \right\rangle - \left\langle \mathbb{1} - \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \right\rangle \\ &\quad = 2\left\langle \Pi_n, \sigma_{a_1 \cdots a_n} \right\rangle + \left\langle \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \right\rangle - 2. \end{aligned} \tag{8.109}$$

Averaging over all choices of $a_1 \cdots a_n \in \Sigma^n$, with each a_k being selected independently according to the probability vector p, one has that

$$\begin{aligned} &\sum_{a_1 \cdots a_n \in \Sigma^n} p(a_1) \cdots p(a_n) \big\langle \Pi_n \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \big\rangle \\ &\quad \geq 2 \big\langle \Pi_n, \sigma^{\otimes n} \big\rangle + \sum_{a_1 \cdots a_n \in \Sigma^n} p(a_1) \cdots p(a_n) \big\langle \Lambda_{a_1 \cdots a_n}, \sigma_{a_1 \cdots a_n} \big\rangle - 2. \end{aligned} \tag{8.110}$$

The right-hand side of the expression (8.110) approaches 1 in the limit as n goes to infinity by Propositions 5.45 and 8.21, from which it follows that

$$\sum_{a_1 \cdots a_n \in \Sigma^n} p(a_1) \cdots p(a_n) \big\langle \mathbb{1} - \Pi_n \Lambda_{a_1 \cdots a_n} \Pi_n, \sigma_{a_1 \cdots a_n} \big\rangle < \frac{\delta}{8} \tag{8.111}$$

for all but finitely many choices of a positive integer n. For any n for which the inequality (8.111) holds, and for a random selection of $g : \Gamma^{m+1} \to \Sigma^n$ as described above, it therefore holds that the expected value of the expression

$$2 \big\langle \mathbb{1} - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n, \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle \tag{8.112}$$

is at most $\delta/4$ for an arbitrary choice of $b_1 \cdots b_{m+1}$, and therefore the same bound holds for a uniformly selected binary string $b_1 \cdots b_{m+1} \in \Gamma^{m+1}$.

The second term in the expression (8.106) will be considered next. It may first be observed that

$$Q - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n = \sum_{\substack{c_1 \cdots c_{m+1} \in \Gamma^{m+1} \\ c_1 \cdots c_{m+1} \neq b_1 \cdots b_{m+1}}} \Pi_n \Lambda_{g(c_1 \cdots c_{m+1})} \Pi_n, \tag{8.113}$$

so that

$$\begin{aligned} &\big\langle Q - \Pi_n \Lambda_{g(b_1 \cdots b_{m+1})} \Pi_n, \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle \\ &= \sum_{\substack{c_1 \cdots c_{m+1} \in \Gamma^{m+1} \\ c_1 \cdots c_{m+1} \neq b_1 \cdots b_{m+1}}} \big\langle \Pi_n \Lambda_{g(c_1 \cdots c_{m+1})} \Pi_n, \sigma_{g(b_1 \cdots b_{m+1})} \big\rangle. \end{aligned} \tag{8.114}$$

The value of the function g on each input string is chosen independently according to the probability vector $p^{\otimes n}$, so there is no correlation between $g(b_1 \cdots b_{m+1})$ and $g(c_1 \cdots c_{m+1})$ for $b_1 \cdots b_{m+1} \neq c_1 \cdots c_{m+1}$. It follows that the expected value of the above expression is given by

$$\big(2^{m+1} - 1\big) \sum_{a_1 \cdots a_n \in \Sigma^n} p(a_1) \cdots p(a_n) \big\langle \Lambda_{a_1 \cdots a_n}, \Pi_n \sigma^{\otimes n} \Pi_n \big\rangle. \tag{8.115}$$

By Proposition 8.25 it holds that

$$\sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n)\operatorname{Tr}\bigl(\Lambda_{a_1\cdots a_n}\bigr) \leq 2^{n(\beta+\varepsilon)} \tag{8.116}$$

for

$$\beta = \sum_{a\in\Sigma} p(a)\,\mathrm{H}(\sigma_a), \tag{8.117}$$

and by the definition of Π_n one has that

$$\lambda_1\bigl(\Pi_n\sigma^{\otimes n}\Pi_n\bigr) \leq 2^{-n(\mathrm{H}(\sigma)-\varepsilon)}. \tag{8.118}$$

It follows that

$$\begin{aligned}\bigl(2^{m+1}-1\bigr)\sum_{a_1\cdots a_n\in\Sigma^n} p(a_1)\cdots p(a_n)\bigl\langle\Lambda_{a_1\cdots a_n},\Pi_n\sigma^{\otimes n}\Pi_n\bigr\rangle \\ \leq 2^{m+1-n(\chi(\eta)-2\varepsilon)},\end{aligned} \tag{8.119}$$

so that the expected value of the second term in the expression (8.106) is upper-bounded by

$$2^{m-n(\chi(\eta)-2\varepsilon)+3}. \tag{8.120}$$

Now assume that $m \leq \alpha n$. For $g : \Gamma^{m+1} \to \Sigma^n$ chosen according to the distribution specified earlier and $b_1\cdots b_{m+1} \in \Gamma^{m+1}$ chosen uniformly, one has that the expected value of the error probability (8.104) is at most

$$\frac{\delta}{4} + 2^{\alpha n-n(\chi(\eta)-2\varepsilon)+3} \leq \frac{\delta}{4} + 2^{-\varepsilon n+3} \tag{8.121}$$

for all but finitely many choices of n. As

$$2^{-\varepsilon n} < \frac{\delta}{32} \tag{8.122}$$

for all sufficiently large n, it follows that the expected value of the error probability (8.104) is smaller than $\delta/2$ for all but finitely many choices of n. For all but finitely many choices of n, there must therefore exist at least one choice of a function $g : \Gamma^{m+1} \to \Sigma^m$ such that, for μ being the measurement associated with g, it holds that

$$\frac{1}{2^{m+1}}\sum_{b_1\cdots b_{m+1}\in\Gamma^{m+1}} \bigl\langle\mathbb{1}-\mu(b_1\cdots b_{m+1}),\sigma_{g(b_1\cdots b_{m+1})}\bigr\rangle < \frac{\delta}{2}. \tag{8.123}$$

Finally, for a given choice of n, $m \leq \alpha n$, g, and μ for which the bound (8.123) holds, consider the set

$$B = \left\{ b_1 \cdots b_{m+1} \in \Gamma^{m+1} \,:\, \left\langle \mathbb{1} - \mu(b_1 \cdots b_{m+1}), \sigma_{g(b_1 \cdots b_{m+1})} \right\rangle \geq \delta \right\} \tag{8.124}$$

of all strings whose encodings incur a decoding error with probability at least δ. It holds that

$$\frac{\delta |B|}{2^{m+1}} < \frac{\delta}{2}, \tag{8.125}$$

and therefore $|B| \leq 2^m$. By defining a function $f : \Gamma^m \to \Sigma^n$ as $f = gh$ for any one-to-one function $h : \Gamma^m \to \Gamma^{m+1} \backslash B$, one has that

$$\left\langle \mu(b_1 \cdots b_m), \sigma_{f(b_1 \cdots b_m)} \right\rangle > 1 - \delta \tag{8.126}$$

for every choice of $b_1 \cdots b_m \in \Gamma^m$, which completes the proof. □

Statement and proof of the Holevo–Schumacher–Westmoreland theorem

The Holevo–Schumacher–Westmoreland theorem will now be stated, and proved through the use of Theorem 8.29.

Theorem 8.30 (Holevo–Schumacher–Westmoreland theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that the classical capacity of Φ is equal to its regularized Holevo capacity:*

$$\mathrm{C}(\Phi) = \lim_{n \to \infty} \frac{\chi\left(\Phi^{\otimes n}\right)}{n}. \tag{8.127}$$

Proof. It will first be observed that the inequality

$$\mathrm{C}(\Phi) \geq \chi(\Phi) \tag{8.128}$$

follows from Theorem 8.29. Let Σ be any alphabet, let

$$\{\rho_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}) \tag{8.129}$$

be a collection of states, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, and define an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ as

$$\eta(a) = p(a)\rho_a \tag{8.130}$$

for each $a \in \Sigma$. Also set $\mathcal{Z} = \mathbb{C}^{\Gamma}$ for $\Gamma = \{0,1\}$ being the binary alphabet. As η is an arbitrarily chosen ensemble, the inequality (8.128) will follow from a demonstration that every positive real number less than $\chi(\Phi(\eta))$ is an achievable rate for classical information transmission through Φ.

Fix any choice of $\alpha > 0$ satisfying

$$\alpha < \chi(\Phi(\eta)), \tag{8.131}$$

and define $\sigma_a = \Phi(\rho_a)$ for each $a \in \Sigma$. By Theorem 8.29, the following statement holds: for every positive real number $\varepsilon > 0$, for all but finitely many choices of a positive integer n, and for all positive integers $m \leq \alpha n$, there exist a classical-to-quantum product state channel code (f, μ) for the collection

$$\{\sigma_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y}) \tag{8.132}$$

for which the probability of an error is strictly less than $\varepsilon/2$ on every binary string of length m.

With this fact in mind, for any fixed choice of positive integers n and m for which $m \leq \alpha n$, one may define encoding and decoding channels

$$\Xi_{\mathrm{E}} \in \mathrm{C}\big(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n}\big) \quad \text{and} \quad \Xi_{\mathrm{D}} \in \mathrm{C}\big(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m}\big) \tag{8.133}$$

as follows:

$$\begin{aligned} \Xi_{\mathrm{E}}(Z) &= \sum_{b_1 \cdots b_m \in \Gamma^m} \big\langle E_{b_1 \cdots b_m, b_1 \cdots b_m}, Z \big\rangle \rho_{f(b_1 \cdots b_m)}, \\ \Xi_{\mathrm{D}}(Y) &= \sum_{b_1 \cdots b_m \in \Gamma^m} \big\langle \mu(b_1 \cdots b_m), Y \big\rangle E_{b_1 \cdots b_m, b_1 \cdots b_m}, \end{aligned} \tag{8.134}$$

for all $Z \in \mathrm{L}(\mathcal{Z}^{\otimes m})$ and $Y \in \mathrm{L}(\mathcal{Y}^{\otimes n})$. For n being sufficiently large, one has that

$$\big\langle E_{b_1 \cdots b_m, b_1 \cdots b_m}, \big(\Xi_{\mathrm{D}} \Phi^{\otimes n} \Xi_{\mathrm{E}}\big)\big(E_{b_1 \cdots b_m, b_1 \cdots b_m}\big) \big\rangle > 1 - \frac{\varepsilon}{2} \tag{8.135}$$

for every $b_1 \cdots b_m \in \Gamma^m$. As Ξ_{E} is a classical-to-quantum channel and Ξ_{D} is quantum-to-classical, it follows that $\Xi_{\mathrm{D}} \Phi^{\otimes n} \Xi_{\mathrm{E}}$ is a ε-approximation to the completely dephasing channel $\Delta^{\otimes m} \in \mathrm{C}(\mathcal{Z}^{\otimes m})$. It has been demonstrated that α is an achievable rate for classical information transmission through Φ.

Following the same reasoning, except replacing the channel Φ by the channel $\Phi^{\otimes n}$, one finds that

$$\chi(\Phi^{\otimes n}) \leq \mathrm{C}(\Phi^{\otimes n}) = n\, \mathrm{C}(\Phi) \tag{8.136}$$

for every positive integer n, and therefore

$$\mathrm{C}(\Phi) \geq \lim_{n\to\infty} \frac{\chi\bigl(\Phi^{\otimes n}\bigr)}{n}. \tag{8.137}$$

It remains to prove that the classical capacity of Φ is no larger than its regularized Holevo capacity. There is nothing to prove if $\mathrm{C}(\Phi) = 0$, so it will be assumed that $\mathrm{C}(\Phi) > 0$. Suppose that $\alpha > 0$ is an achievable rate for classical information transmission through Φ, and let $\varepsilon > 0$ be chosen arbitrarily. It must therefore hold, for all but finitely many positive integers n, and for all $m \leq \alpha n$, that $\Phi^{\otimes n}$ emulates an ε-approximation to the completely dephasing channel $\Delta^{\otimes m} \in \mathrm{C}(\mathcal{Z}^{\otimes m})$.

Let n be any positive integer for which this property holds and for which $\lfloor \alpha n \rfloor \geq 2$, and let $m = \lfloor \alpha n \rfloor$. The situation in which a sender generates a binary string of length m, uniformly at random, and transmits this string through the ε-approximation to $\Delta^{\otimes m}$ emulated by $\Phi^{\otimes n}$ will be considered. Let X and Z be classical registers both having state set Γ^m, where X is a register representing the randomly generated string selected by the sender and Z is a register representing the string obtained by the receiver when a copy of the string stored in X is transmitted through the ε-approximation to $\Delta^{\otimes m}$ emulated by $\Phi^{\otimes n}$. As $\Phi^{\otimes n}$ emulates an ε-approximation to $\Delta^{\otimes m}$, there must exist a collection of states

$$\bigl\{\rho_{b_1\cdots b_m} \,:\, b_1\cdots b_m \in \Gamma^m\bigr\} \subseteq \mathrm{D}\bigl(\mathcal{X}^{\otimes n}\bigr) \tag{8.138}$$

along with a measurement $\mu : \Gamma^m \to \mathrm{Pos}\bigl(\mathcal{Y}^{\otimes n}\bigr)$ such that

$$\bigl\langle \mu(b_1\cdots b_m), \Phi^{\otimes n}(\rho_{b_1\cdots b_m})\bigr\rangle \geq 1 - \frac{\varepsilon}{2} \tag{8.139}$$

for every binary string $b_1\cdots b_m \in \Gamma^m$. With respect to the probability vector $p \in \mathcal{P}(\Gamma^m \times \Gamma^m)$ defined as

$$p(b_1\cdots b_m, c_1\cdots c_m) = \frac{1}{2^m}\bigl\langle \mu(c_1\cdots c_m), \Phi^{\otimes n}(\rho_{b_1\cdots b_m})\bigr\rangle, \tag{8.140}$$

which represents the probabilistic state of (X,Z) suggested above, it follows from Holevo's theorem (Theorem 5.52) that

$$\mathrm{I}(\mathsf{X}:\mathsf{Z}) \leq \chi(\Phi^{\otimes n}(\eta)), \tag{8.141}$$

where $\eta : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes n})$ is the ensemble defined as

$$\eta(b_1 \cdots b_m) = \frac{1}{2^m} \rho_{b_1 \cdots b_m} \tag{8.142}$$

for each $b_1 \cdots b_m \in \Gamma^m$.

A lower-bound on the mutual information $\mathrm{I}(\mathsf{X} : \mathsf{Z})$ may be derived as follows. The distribution represented by the marginal probability vector $p[\mathsf{X}]$ is uniform, and therefore $\mathrm{H}(p[\mathsf{X}]) = m$. By (8.139), each entry of the probability vector $p[\mathsf{Z}]$ is lower-bounded by $(1 - \varepsilon/2)2^{-m}$, so the lower bound

$$\mathrm{H}(p[\mathsf{Z}]) \geq \left(1 - \frac{\varepsilon}{2}\right) m \tag{8.143}$$

follows by the concavity of the Shannon entropy function (Proposition 5.5). Finally, again making use of (8.139), one may conclude that

$$\begin{aligned} \mathrm{H}(p) &\leq -\left(1 - \frac{\varepsilon}{2}\right) \log\left(\frac{1 - \varepsilon/2}{2^m}\right) - \frac{\varepsilon}{2} \log\left(\frac{\varepsilon/2}{2^{2m} - 2^m}\right) \\ &< \left(1 + \frac{\varepsilon}{2}\right) m + \mathrm{H}\left(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2}\right). \end{aligned} \tag{8.144}$$

It follows that

$$\begin{aligned} \chi\big(\Phi^{\otimes n}\big) \geq \mathrm{I}(\mathsf{X} : \mathsf{Z}) &= \mathrm{H}(p[\mathsf{X}]) + \mathrm{H}(p[\mathsf{Z}]) - \mathrm{H}(p) \\ &\geq (1 - \varepsilon) m - \mathrm{H}\left(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2}\right). \end{aligned} \tag{8.145}$$

Now, given that $\varepsilon > 0$ was chosen arbitrarily, it follows that

$$\chi\big(\Phi^{\otimes n}\big) \geq m = \lfloor \alpha n \rfloor, \tag{8.146}$$

for all but finitely many positive integers n. One therefore has that

$$\alpha \leq \lim_{n \to \infty} \frac{\chi\big(\Phi^{\otimes n}\big)}{n}. \tag{8.147}$$

As $\mathrm{C}(\Phi)$ is equal to the supremum over all achievable rates α for classical information transmission through Φ, it follows that

$$\mathrm{C}(\Phi) \leq \lim_{n \to \infty} \frac{\chi\big(\Phi^{\otimes n}\big)}{n}, \tag{8.148}$$

as required. □

8.1.3 The entanglement-assisted classical capacity theorem

This section is focused on the *entanglement-assisted classical capacity theorem*, which characterizes the entanglement-assisted classical capacity of a given channel. This theorem stands out among the capacity theorems presented in the present chapter, as no regularization is required by the characterization it provides.

Holevo–Schumacher–Westmoreland with entanglement assistance

A preliminary step toward the proof of the entanglement-assisted classical capacity theorem is the observation that, when both the classical capacity and Holevo capacity are replaced by their entanglement-assisted forms, an analogous statement to the Holevo–Schumacher–Westmoreland theorem is true.

Theorem 8.31. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The entanglement-assisted classical capacity of Φ equals the regularized entanglement-assisted Holevo capacity of Φ:*

$$\mathrm{C}_{\mathrm{E}}(\Phi) = \lim_{n \to \infty} \frac{\chi_{\mathrm{E}}\bigl(\Phi^{\otimes n}\bigr)}{n}. \tag{8.149}$$

Proof. The theorem is proved in essentially the same way as the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30), with each step being modified to allow for the possibility of entanglement assistance.

In greater detail, let $\mathcal{W}$ be a complex Euclidean space and let η be an ensemble of the form

$$\eta : \Sigma \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{W}) \tag{8.150}$$

that is constant with respect to $\mathcal{W}$. By Proposition 8.12, one may choose a complex Euclidean space $\mathcal{V}$, a state $\xi \in \mathrm{D}(\mathcal{V} \otimes \mathcal{W})$, a probability vector $p \in \mathcal{P}(\Sigma)$, and a collection of channels

$$\{\Psi_a \,:\, a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{V}, \mathcal{X}) \tag{8.151}$$

such that

$$\eta(a) = p(a)\bigl(\Psi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\bigr)(\xi) \tag{8.152}$$

for every $a \in \Sigma$. It will be proved that every positive real number α with

$$\alpha < \chi\Bigl(\bigl(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\bigr)(\eta)\Bigr) \tag{8.153}$$

is an achievable rate for entanglement-assisted classical information transmission through Φ. Let

$$\sigma_a = \big(\Phi\Psi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\xi) \tag{8.154}$$

for each $a \in \Sigma$.

By Theorem 8.29 it follows, for every positive real number $\varepsilon > 0$, for all but finitely many choices of a positive integer n, and for all positive integers $m \leq \alpha n$, that there exist a classical-to-quantum product state channel code (f, μ) for the collection

$$\{\sigma_a \,:\, a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{Y} \otimes \mathcal{W}) \tag{8.155}$$

having error bounded by $\varepsilon/2$. Assume hereafter that such a code, taking the form

$$f : \Gamma^m \to \Sigma^n \quad \text{and} \quad \mu : \Gamma^m \to \mathrm{Pos}\big((\mathcal{Y} \otimes \mathcal{W})^{\otimes n}\big) \tag{8.156}$$

has been selected.

It will be proved that the channel $\Phi^{\otimes n}$ emulates a ε-approximation to the completely dephasing channel $\Delta^{\otimes m} \in \mathrm{C}(\mathcal{Z}^{\otimes m})$ with the assistance of entanglement. The entangled state to be used to assist this emulation is

$$V\xi^{\otimes n}V^* \in \mathrm{D}\big(\mathcal{V}^{\otimes n} \otimes \mathcal{W}^{\otimes n}\big), \tag{8.157}$$

where $V \in \mathrm{U}\big((\mathcal{V} \otimes \mathcal{W})^{\otimes n}, \mathcal{V}^{\otimes n} \otimes \mathcal{W}^{\otimes n}\big)$ represents a permutation of tensor factors:

$$\begin{aligned} &V\big((v_1 \otimes w_1) \otimes \cdots \otimes (v_n \otimes w_n)\big) \\ &\quad = (v_1 \otimes \cdots \otimes v_n) \otimes (w_1 \otimes \cdots \otimes w_n) \end{aligned} \tag{8.158}$$

for every choice of vectors $v_1, \ldots, v_n \in \mathcal{V}$ and $w_1, \ldots, w_n \in \mathcal{W}$. Define an encoding channel $\Xi_{\mathrm{E}} \in \mathrm{C}\big(\mathcal{Z}^{\otimes m} \otimes \mathcal{V}^{\otimes n}, \mathcal{X}^{\otimes n}\big)$ as

$$\Xi_{\mathrm{E}} = \sum_{b_1 \cdots b_m \in \Gamma^m} \Theta_{b_1 \cdots b_m} \otimes \Psi_{f(b_1 \cdots b_m)}, \tag{8.159}$$

where one defines

$$\Psi_{a_1 \cdots a_n} = \Psi_{a_1} \otimes \cdots \otimes \Psi_{a_n} \tag{8.160}$$

for each $a_1 \cdots a_n \in \Sigma^n$, and where $\Theta_{b_1 \cdots b_m} \in \mathrm{CP}(\mathcal{Z}^{\otimes m}, \mathbb{C})$ is given by

$$\Theta_{b_1 \cdots b_m}(Z) = Z(b_1 \cdots b_m, b_1 \cdots b_m) \tag{8.161}$$

for every $Z \in \mathrm{L}(\mathcal{Z}^{\otimes m})$. Described in words, the encoding map Ξ_{E} takes as input a compound register $(\mathsf{Z}_1, \ldots, \mathsf{Z}_m, \mathsf{V}_1, \ldots, \mathsf{V}_n)$, measures $(\mathsf{Z}_1, \ldots, \mathsf{Z}_m)$ with respect to the standard basis measurement, and applies the channel $\Psi_{f(b_1 \cdots b_m)}$ to $(\mathsf{V}_1, \ldots, \mathsf{V}_n)$ for $b_1 \cdots b_m$ being the string obtained from the standard basis measurement on $(\mathsf{Z}_1, \ldots, \mathsf{Z}_m)$. Define a decoding channel $\Xi_{\mathrm{D}} \in \mathrm{C}\big(\mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n}, \mathcal{Z}^{\otimes m}\big)$ as

$$\Xi_{\mathrm{D}}(Y) = \sum_{b_1 \cdots b_m \in \Gamma^m} \big\langle W \mu(b_1 \cdots b_m) W^*, Y \big\rangle \, E_{b_1 \cdots b_m, b_1 \cdots b_m} \tag{8.162}$$

for all $Y \in \mathrm{L}(\mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n})$, where $W \in \mathrm{U}\big((\mathcal{Y} \otimes \mathcal{W})^{\otimes n}, \mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n}\big)$ is an isometry representing a permutation of tensor factors that is similar to V, but with $\mathcal{V}$ replaced by $\mathcal{Y}$:

$$\begin{aligned} &W\big((y_1 \otimes w_1) \otimes \cdots \otimes (y_n \otimes w_n)\big) \\ &\quad = (y_1 \otimes \cdots \otimes y_n) \otimes (w_1 \otimes \cdots \otimes w_n) \end{aligned} \tag{8.163}$$

for all choices of vectors $y_1, \ldots, y_n \in \mathcal{Y}$ and $w_1, \ldots, w_n \in \mathcal{W}$.

Now, let $\Psi \in \mathrm{C}(\mathcal{Z}^{\otimes m})$ denote the channel that has been emulated with the assistance of entanglement by the above construction:

$$\Psi(Z) = \big(\Xi_{\mathrm{D}}\big(\Phi^{\otimes n} \Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}^{\otimes n}\big)\big)\big(Z \otimes V \xi^{\otimes n} V^*\big) \tag{8.164}$$

for every $Z \in \mathrm{L}(\mathcal{Z}^{\otimes m})$. For every binary string $b_1 \cdots b_m \in \Gamma^m$, it holds that

$$\big(\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}^{\otimes n}\big)\big(E_{b_1 \cdots b_m, b_1 \cdots b_m} \otimes V \xi^{\otimes n} V^*\big) = W \rho_{f(b_1 \cdots b_m)} W^*, \tag{8.165}$$

from which it follows that

$$\big\langle E_{b_1 \cdots b_m, b_1 \cdots b_m}, \Psi(E_{b_1 \cdots b_m, b_1 \cdots b_m}) \big\rangle \geq 1 - \frac{\varepsilon}{2}. \tag{8.166}$$

As $\Psi = \Delta^{\otimes m} \Psi \Delta^{\otimes m}$, it follows that Ψ is a ε-approximation to $\Delta^{\otimes m}$.

As $\varepsilon > 0$ has been chosen arbitrarily, and the analysis above may be considered for all but finitely many n and all $m \leq \alpha n$, one may conclude that α is an achievable rate for entanglement-assisted classical information transmission through Φ. Given that the alphabet Σ, the complex Euclidean space $\mathcal{W}$, and the ensemble η were chosen arbitrarily, subject to η being constant with respect to $\mathcal{W}$, it follows that

$$\chi_{\mathrm{E}}(\Phi) \leq \mathrm{C}_{\mathrm{E}}(\Phi). \tag{8.167}$$

Applying the same argument to the channel $\Phi^{\otimes n}$ in place of Φ, for any choice of a positive integer n, yields

$$\chi_{\mathrm{E}}(\Phi^{\otimes n}) \leq \mathrm{C}_{\mathrm{E}}(\Phi^{\otimes n}) = n\,\mathrm{C}_{\mathrm{E}}(\Phi), \tag{8.168}$$

and therefore

$$\mathrm{C}_{\mathrm{E}}(\Phi) \geq \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\bigl(\Phi^{\otimes n}\bigr)}{n}. \tag{8.169}$$

It remains to prove that the entanglement-assisted classical capacity of Φ cannot exceed its regularized entanglement-assisted Holevo capacity. As in the proof of Theorem 8.30, it may be assumed that $\mathrm{C}_{\mathrm{E}}(\Phi) > 0$, and it suffices to consider the situation in which a sender transmits a uniformly generated binary string of length m to a receiver.

Suppose $\alpha > 0$ is an achievable rate for entanglement-assisted classical information transmission through Φ, and let $\varepsilon > 0$ be chosen arbitrarily. It must therefore hold, for all but finitely many positive integers n, and for all $m \leq \alpha n$, that $\Phi^{\otimes n}$ emulates an ε-approximation to $\Delta^{\otimes m}$ with the assistance of entanglement. Let n be an arbitrarily chosen positive integer for which this property holds and for which $\lfloor \alpha n \rfloor \geq 2$, and let $m = \lfloor \alpha n \rfloor$.

By the assumption that $\Phi^{\otimes n}$ emulates an ε-approximation to $\Delta^{\otimes m}$ with the assistance of entanglement, one may conclude that there exists a choice of complex Euclidean spaces $\mathcal{V}$ and $\mathcal{W}$, a state $\xi \in \mathrm{D}(\mathcal{V} \otimes \mathcal{W})$, a collection of channels

$$\bigl\{\Psi_{b_1\cdots b_m} \,:\, b_1 \cdots b_m \in \Gamma^m\bigr\} \subseteq \mathrm{C}\bigl(\mathcal{V}, \mathcal{X}^{\otimes n}\bigr), \tag{8.170}$$

and a measurement $\mu : \Gamma^m \to \mathrm{Pos}(\mathcal{Y}^{\otimes n} \otimes \mathcal{W})$, such that

$$\Bigl\langle \mu(b_1 \cdots b_m), \bigl(\Phi^{\otimes n}\Psi_{b_1\cdots b_m} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\bigr)(\xi) \Bigr\rangle \geq 1 - \frac{\varepsilon}{2} \tag{8.171}$$

for every string $b_1 \cdots b_m \in \Gamma^m$.

Let X and Z be classical registers both having state set Γ^m, where X is a register representing the randomly generated string selected by the sender and Z is a register representing the string obtained by the receiver when a copy of the string stored in X is transmitted through the ε-approximation to $\Delta^{\otimes m}$ emulated by $\Phi^{\otimes n}$ with the assistance of entanglement. With respect to the probability vector $p \in \mathcal{P}(\Gamma^m \times \Gamma^m)$ defined as

$$\begin{aligned} &p(b_1 \cdots b_m, c_1 \cdots c_m) \\ &\quad = \frac{1}{2^m} \Bigl\langle \mu(c_1 \cdots c_m), \bigl(\Phi^{\otimes n}\Psi_{b_1\cdots b_m} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\bigr)(\xi) \Bigr\rangle, \end{aligned} \tag{8.172}$$

representing a probabilistic state of (X, Z), it follows from Holevo's theorem (Theorem 5.52) that

$$\mathrm{I}(\mathsf{X} : \mathsf{Z}) \leq \chi\big(\big(\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\eta)\big), \tag{8.173}$$

for $\eta : \Gamma^m \to \mathrm{Pos}(\mathcal{X}^{\otimes n} \otimes \mathcal{W})$ being the ensemble defined as

$$\eta(b_1 \cdots b_m) = \frac{1}{2^m}\big(\Psi_{b_1 \cdots b_m} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\xi). \tag{8.174}$$

The same lower-bound on the quantity $\mathrm{I}(\mathsf{X} : \mathsf{Z})$ derived in the proof of Theorem 8.30 holds in the present case, from which it follows that

$$\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big) \geq \mathrm{I}(\mathsf{X} : \mathsf{Z}) \geq (1 - \varepsilon)m - \mathrm{H}\Big(1 - \frac{\varepsilon}{2}, \frac{\varepsilon}{2}\Big). \tag{8.175}$$

Given that $\varepsilon > 0$ was chosen arbitrarily, one has that

$$\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big) \geq m = \lfloor \alpha n \rfloor, \tag{8.176}$$

for all but finitely many positive integers n. Consequently,

$$\alpha \leq \lim_{n \to \infty} \frac{\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big)}{n}. \tag{8.177}$$

As $\mathrm{C}_{\mathrm{E}}(\Phi)$ is defined as the supremum over all achievable rates α for classical information transmission through Φ with the assistance of entanglement, it follows that

$$\mathrm{C}_{\mathrm{E}}(\Phi) \leq \lim_{n \to \infty} \frac{\chi_{\mathrm{E}}\big(\Phi^{\otimes n}\big)}{n}, \tag{8.178}$$

which completes the proof. □

Strongly typical strings and projections

The proof of the entanglement-assisted classical capacity theorem that is presented in this book will make use of a notion of typicality, known as *strong typicality*, that differs from the standard notion discussed previously in Section 5.3.1. True to its name, strong typicality is the more restrictive of the two notions; every strongly typical string will necessarily be a typical string, up to a simple change of parameters, while some typical strings are not strongly typical.

Similar to the standard notion of typicality, one may define an *ε-strongly typical subspace* with respect to a spectral decomposition of a given state. Unlike the standard typical subspace, the strongly typical subspace may not be uniquely determined by a given state—when the spectral decomposition is not unique, the strongly typical subspace may depend on the particular spectral decomposition with respect to which it was defined. Despite this apparent drawback, the notion of an ε-strongly typical subspace will prove to be an important concept in proving the entanglement-assisted classical capacity theorem.

The definition of strong-typicality to follow uses the following notation, for which it is to be assumed that Σ is an alphabet and n is a positive integer. For every string $a_1 \cdots a_n \in \Sigma^n$ and symbol $a \in \Sigma$, one writes

$$N(a \,|\, a_1 \cdots a_n) = \big|\{k \in \{1, \ldots, n\} \,:\, a_k = a\}\big|, \tag{8.179}$$

which is the number of times the symbol a occurs in the string $a_1 \cdots a_n$.

Definition 8.32. Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\varepsilon > 0$ be a positive real number. A string $a_1 \cdots a_n \in \Sigma^n$ is said to be *ε-strongly typical with respect to p* if and only if

$$\left| \frac{N(a \,|\, a_1 \cdots a_n)}{n} - p(a) \right| \leq p(a)\varepsilon \tag{8.180}$$

for every $a \in \Sigma$. The set of all ε-strongly typical strings of length n with respect to p is denoted $S_{n,\varepsilon}(p)$ (or by $S_{n,\varepsilon}$ when p is implicit and can safely be omitted).

The average behavior of a nonnegative real-valued function defined on the individual symbols of a strongly typical string may be analyzed using the following elementary proposition.

Proposition 8.33. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, let $\varepsilon > 0$ be a positive real number, let $a_1 \cdots a_n \in S_{n,\varepsilon}(p)$ be an ε-strongly typical string with respect to p, and let $\phi : \Sigma \to [0, \infty)$ be a nonnegative real-valued function. It holds that*

$$\left| \frac{\phi(a_1) + \cdots + \phi(a_n)}{n} - \sum_{a \in \Sigma} p(a)\phi(a) \right| \leq \varepsilon \sum_{a \in \Sigma} p(a)\phi(a). \tag{8.181}$$

Proof. The inequality (8.181) follows from the definition of strong typicality together with the triangle inequality:

$$\begin{aligned}
&\left|\frac{\phi(a_1) + \cdots + \phi(a_n)}{n} - \sum_{a\in\Sigma} p(a)\phi(a)\right| \\
&\qquad = \left|\sum_{a\in\Sigma}\left(\frac{N(a \mid a_1\cdots a_n)\phi(a)}{n} - p(a)\phi(a)\right)\right| \\
&\qquad \leq \sum_{a\in\Sigma}\phi(a)\left|\frac{N(a \mid a_1\cdots a_n)}{n} - p(a)\right| \\
&\qquad \leq \varepsilon \sum_{a\in\Sigma} p(a)\phi(a),
\end{aligned} \tag{8.182}$$

as required. □

As a corollary to Proposition 8.33, one has that every ε-strongly typical string, with respect to a given probability vector p, is necessarily δ-typical for every choice of $\delta > \varepsilon\,\mathrm{H}(p)$.

Corollary 8.34. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, let $\varepsilon > 0$ be a positive real number, and let $a_1\cdots a_n \in S_{n,\varepsilon}(p)$ be an ε-strongly typical string with respect to p. It holds that*

$$2^{-n(1+\varepsilon)\,\mathrm{H}(p)} \leq p(a_1)\cdots p(a_n) \leq 2^{-n(1-\varepsilon)\,\mathrm{H}(p)}. \tag{8.183}$$

Proof. Define a function $\phi : \Sigma \to [0,\infty)$ as

$$\phi(a) = \begin{cases} -\log(p(a)) & \text{if } p(a) \neq 0 \\ 0 & \text{if } p(a) = 0. \end{cases} \tag{8.184}$$

With respect to this function, the implication provided by Proposition 8.33 is equivalent to (8.183). □

Strings that are obtained by independently selecting symbols at random according to a given probability vector are likely to be not only typical, but strongly typical, with the probability of strong typicality increasing with string length. The following lemma establishes a quantitative bound on this probability.

Lemma 8.35. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\varepsilon > 0$ be a positive real number. It holds that*

$$\sum_{a_1 \cdots a_n \in S_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) \geq 1 - \zeta_{n,\varepsilon}(p) \tag{8.185}$$

for

$$\zeta_{n,\varepsilon}(p) = 2 \sum_{a \in \operatorname{supp}(p)} \exp\bigl(-n\varepsilon^2 p(a)^2\bigr). \tag{8.186}$$

Proof. Suppose first that $a \in \Sigma$ is fixed, and consider the probability that a string $a_1 \cdots a_n \in \Sigma^n$, randomly selected according to the probability vector $p^{\otimes n}$, satisfies

$$\left| \frac{N(a \,|\, a_1 \cdots a_n)}{n} - p(a) \right| > p(a)\varepsilon. \tag{8.187}$$

To bound this probability, one may define $X_1, \ldots, X_n$ to be independent and identically distributed random variables, taking value 1 with probability $p(a)$ and value 0 otherwise, so that the probability of the event (8.187) is equal to

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\varepsilon \right). \tag{8.188}$$

If it is the case that $p(a) > 0$, then Hoeffding's inequality (Theorem 1.20) implies that

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\varepsilon \right) \leq 2\exp\bigl(-n\varepsilon^2 p(a)^2\bigr), \tag{8.189}$$

while it holds that

$$\Pr\left(\left| \frac{X_1 + \cdots + X_n}{n} - p(a) \right| > p(a)\varepsilon \right) = 0 \tag{8.190}$$

in case $p(a) = 0$. The lemma follows from the union bound. □

The next proposition establishes upper and lower bounds on the number of strings in an ε-strongly typical set for a given length.

Proposition 8.36. *Let Σ be an alphabet, let $p \in \mathcal{P}(\Sigma)$ be a probability vector, let n be a positive integer, and let $\varepsilon > 0$ be a positive real number. It holds that*

$$\bigl(1 - \zeta_{n,\varepsilon}(p)\bigr)\, 2^{n(1-\varepsilon)\,\mathrm{H}(p)} \leq \bigl|S_{n,\varepsilon}(p)\bigr| \leq 2^{n(1+\varepsilon)\,\mathrm{H}(p)}, \tag{8.191}$$

for $\zeta_{n,\varepsilon}(p)$ as defined in Lemma 8.35.

Proof. By Corollary 8.34, one has

$$p(a_1) \cdots p(a_n) \geq 2^{-n(1+\varepsilon)\,\mathrm{H}(p)} \tag{8.192}$$

for every string $a_1 \cdots a_n \in S_{n,\varepsilon}(p)$. Consequently,

$$1 \geq \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) \geq \big|S_{n,\varepsilon}(p)\big| 2^{-n(1+\varepsilon)\,\mathrm{H}(p)}, \tag{8.193}$$

and therefore

$$\big|S_{n,\varepsilon}(p)\big| \leq 2^{n(1+\varepsilon)\,\mathrm{H}(p)}. \tag{8.194}$$

Along similar lines, one has

$$p(a_1) \cdots p(a_n) \leq 2^{-n(1-\varepsilon)\,\mathrm{H}(p)} \tag{8.195}$$

for every string $a_1 \cdots a_n \in S_{n,\varepsilon}(p)$. By Lemma 8.35, it follows that

$$1 - \zeta_{n,\varepsilon}(p) \leq \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}(p)} p(a_1) \cdots p(a_n) \leq \big|S_{n,\varepsilon}(p)\big|\, 2^{-n(1-\varepsilon)\,\mathrm{H}(p)}, \tag{8.196}$$

and therefore

$$\big|S_{n,\varepsilon}(p)\big| \geq \big(1 - \zeta_{n,\varepsilon}(p)\big)\, 2^{n(1-\varepsilon)\,\mathrm{H}(p)}, \tag{8.197}$$

as required. □

Finally, the ε-strongly typical subspaces associated with a given density operator are defined as follows.

Definition 8.37. Let $\mathcal{X}$ be a complex Euclidean space, let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, let $\varepsilon > 0$ be a positive real number, and let n be a positive integer. Also let

$$\rho = \sum_{a \in \Sigma} p(a) x_a x_a^* \tag{8.198}$$

be a spectral decomposition of ρ, for Σ being an alphabet, $p \in \mathcal{P}(\Sigma)$ being a probability vector, and $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ being an orthonormal set of vectors. The *projection operator onto the ε-strongly typical subspace of $\mathcal{X}^{\otimes n}$* with respect to the spectral decomposition (8.198) is defined as

$$\Lambda = \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}(p)} x_{a_1} x_{a_1}^* \otimes \cdots \otimes x_{a_n} x_{a_n}^*. \tag{8.199}$$

With respect to the decomposition (8.198), the *ε-strongly typical subspace of $\mathcal{X}^{\otimes n}$* is defined as the image of Λ.

Example 8.38. Let $\Sigma = \{0,1\}$, let $\mathcal{X} = \mathbb{C}^{\Sigma}$, and let $\rho = \mathbb{1}/2 \in \mathrm{D}(\mathcal{X})$. With respect to the spectral decomposition

$$\rho = \frac{1}{2} e_0 e_0^* + \frac{1}{2} e_1 e_1^*, \tag{8.200}$$

for $n = 2$ and for any choice of $\varepsilon \in (0,1)$, one has that the corresponding projection operator onto the ε-strongly typical subspace is given by

$$\Lambda_0 = E_{0,0} \otimes E_{1,1} + E_{1,1} \otimes E_{0,0}. \tag{8.201}$$

Replacing the spectral decomposition by

$$\rho = \frac{1}{2} x_0 x_0^* + \frac{1}{2} x_1 x_1^*, \tag{8.202}$$

for

$$x_0 = \frac{e_0 + e_1}{\sqrt{2}} \quad \text{and} \quad x_1 = \frac{e_0 - e_1}{\sqrt{2}}, \tag{8.203}$$

one obtains the corresponding projection operator

$$\Lambda_1 = x_0 x_0^* \otimes x_1 x_1^* + x_1 x_1^* \otimes x_0 x_0^* \neq \Lambda_0. \tag{8.204}$$

Two lemmas on the output entropy of channels

The proof of the entanglement-assisted classical capacity theorem appearing at the end of the present section will make use of multiple lemmas. The two lemmas that follow concern the output entropy of channels. The first of these two lemmas will also be used in the next section of the chapter, when proving that the coherent information is a lower bound on the quantum capacity of a channel.

Lemma 8.39. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let $\rho \in \mathrm{D}(\mathcal{X})$ be a density operator, and let $\varepsilon > 0$ be a positive real number. Let*

$$\rho = \sum_{a \in \Sigma} p(a) x_a x_a^* \tag{8.205}$$

be a spectral decomposition of ρ, for Σ being an alphabet, $\{x_a : a \in \Sigma\} \subset \mathcal{X}$ being an orthonormal set, and $p \in \mathcal{P}(\Sigma)$ being a probability vector. For every positive integer n, let $\Lambda_{n,\varepsilon}$ denote the projection operator onto the ε-strongly typical subspace of $\mathcal{X}^{\otimes n}$ with respect to the decomposition (8.205), and let

$$\omega_{n,\varepsilon} = \frac{\Lambda_{n,\varepsilon}}{\mathrm{Tr}(\Lambda_{n,\varepsilon})}. \tag{8.206}$$

It holds that

$$\left| \lim_{n\to\infty} \frac{\mathrm{H}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon})\big)}{n} - \mathrm{H}(\Phi(\rho)) \right| \leq \big(2\,\mathrm{H}(\rho) + \mathrm{H}(\Phi(\rho))\big)\,\varepsilon. \tag{8.207}$$

Proof. It may be verified that the equation

$$\begin{aligned} &\mathrm{H}(\Phi(\rho)) - \frac{1}{n}\,\mathrm{H}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon})\big) \\ &\qquad = \frac{1}{n}\,\mathrm{D}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon}) \,\big\|\, \Phi^{\otimes n}\big(\rho^{\otimes n}\big)\big) \\ &\qquad\quad + \frac{1}{n}\,\mathrm{Tr}\Big(\big(\Phi^{\otimes n}(\omega_{n,\varepsilon}) - \Phi(\rho)^{\otimes n}\big)\log\big(\Phi(\rho)^{\otimes n}\big)\Big) \end{aligned} \tag{8.208}$$

holds for every positive integer n. Bounds on the absolute values of the two terms on the right-hand side of this equation will be established separately.

The first term on the right-hand side of (8.208) is nonnegative, and an upper bound on it may be obtained from the monotonicity of the quantum relative entropy under the action of channels (Theorem 5.38). Specifically, by Corollary 8.34 and Proposition 8.36, one has

$$\begin{aligned} &\frac{1}{n}\,\mathrm{D}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon}) \,\big\|\, \Phi^{\otimes n}(\rho^{\otimes n})\big) \leq \frac{1}{n}\,\mathrm{D}\big(\omega_{n,\varepsilon} \,\big\|\, \rho^{\otimes n}\big) \\ &= -\frac{1}{n}\log(|S_{n,\varepsilon}|) - \frac{1}{n|S_{n,\varepsilon}|} \sum_{a_1\cdots a_n \in S_{n,\varepsilon}} \log(p(a_1)\cdots p(a_n)) \\ &\qquad\qquad \leq 2\varepsilon\,\mathrm{H}(\rho) - \frac{\log(1-\zeta_{n,\varepsilon})}{n} \end{aligned} \tag{8.209}$$

for every positive integer n, where $S_{n,\varepsilon}$ denotes the set of ε-strongly typical strings of length n with respect to p and $\zeta_{n,\varepsilon}$ denotes the quantity defined in Lemma 8.35.

To bound the absolute value of second term on the right-hand side of (8.208), one may first define a function $\phi : \Sigma \to [0,\infty)$ as

$$\phi(a) = \begin{cases} -\mathrm{Tr}(\Phi(x_a x_a^*)\log(\Phi(\rho))) & \text{if } p(a) > 0 \\ 0 & \text{if } p(a) = 0 \end{cases} \tag{8.210}$$

for each $a \in \Sigma$. It is evident from its specification that $\phi(a)$ is nonnegative for each $a \in \Sigma$, and is finite by virtue of the fact that

$$\mathrm{im}(\Phi(x_a x_a^*)) \subseteq \mathrm{im}(\Phi(\rho)) \tag{8.211}$$

for each $a \in \operatorname{supp}(p)$. Using the identity

$$\log\bigl(P^{\otimes n}\bigr) = \sum_{k=1}^{n} \mathbb{1}^{\otimes(k-1)} \otimes \log(P) \otimes \mathbb{1}^{\otimes(n-k)}, \tag{8.212}$$

it may be verified that

$$\begin{aligned} &\operatorname{Tr}\bigl(\Phi^{\otimes n}(\omega_{n,\varepsilon}) \log\bigl(\Phi(\rho)^{\otimes n}\bigr)\bigr) \\ &\quad = -\frac{1}{|S_{n,\varepsilon}|} \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}} (\phi(a_1) + \cdots + \phi(a_n)), \end{aligned} \tag{8.213}$$

and it is evident that

$$\mathrm{H}(\Phi(\rho)) = \sum_{a \in \Sigma} p(a)\phi(a). \tag{8.214}$$

It therefore holds that

$$\begin{aligned} &\left|\frac{1}{n} \operatorname{Tr}\bigl(\bigl(\Phi^{\otimes n}(\omega_{n,\varepsilon}) - \Phi(\rho)^{\otimes n}\bigr) \log\bigl(\Phi(\rho)^{\otimes n}\bigr)\bigr)\right| \\ &\qquad \leq \frac{1}{|S_{n,\varepsilon}|} \sum_{a_1 \cdots a_n \in S_{n,\varepsilon}} \left| \frac{\phi(a_1) + \cdots + \phi(a_n)}{n} - \sum_{a \in \Sigma} p(a)\phi(a) \right| \\ &\qquad \leq \varepsilon \sum_{a \in \Sigma} p(a)\phi(a) \\ &\qquad = \varepsilon\, \mathrm{H}(\Phi(\rho)). \end{aligned} \tag{8.215}$$

Combining the inequalities (8.209) and (8.215), one has

$$\begin{aligned} &\left|\frac{1}{n} \mathrm{H}\bigl(\Phi^{\otimes n}(\omega_{n,\varepsilon})\bigr) - \mathrm{H}(\Phi(\rho))\right| \\ &\qquad \leq 2\varepsilon\, \mathrm{H}(\rho) - \frac{\log(1 - \zeta_{n,\varepsilon})}{n} + \varepsilon\, \mathrm{H}(\Phi(\rho)), \end{aligned} \tag{8.216}$$

from which the lemma follows. □

Lemma 8.40. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The function $f : \mathrm{D}(\mathcal{X}) \to \mathbb{R}$ defined by*

$$f(\rho) = \mathrm{H}(\rho) - \mathrm{H}(\Phi(\rho)) \tag{8.217}$$

is concave.

Proof. Let $\mathcal{Z}$ be any complex Euclidean space, and consider first the function $g : \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z}) \to \mathbb{R}$ defined as

$$g(\sigma) = \mathrm{H}(\sigma) - \mathrm{H}(\mathrm{Tr}_{\mathcal{Z}}(\sigma)) \tag{8.218}$$

for every $\sigma \in \mathrm{D}(\mathcal{Y} \otimes \mathcal{Z})$. An alternative expression for g is

$$g(\sigma) = -\mathrm{D}(\sigma \,\|\, \mathrm{Tr}_{\mathcal{Z}}(\sigma) \otimes \mathbb{1}_{\mathcal{Z}}), \tag{8.219}$$

and the concavity of g therefore follows from the joint convexity of quantum relative entropy (Corollary 5.36).

For a suitable choice of a complex Euclidean space $\mathcal{Z}$, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be an isometry that yields a Stinespring representation of Φ:

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \tag{8.220}$$

for every $X \in \mathrm{L}(\mathcal{X})$. The function f is given by $f(\rho) = g(A\rho A^*)$ for every $\rho \in \mathrm{D}(\mathcal{X})$, and therefore the concavity of g implies that f is concave as well. □

An additivity lemma concerning the coherent information

The next lemma that will be used in the proof of the entanglement-assisted capacity theorem is the following lemma, which states that the quantity

$$\max_{\sigma \in \mathrm{D}(\mathcal{X})} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma\,;\Phi)\big), \tag{8.221}$$

defined for each channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, is additive with respect to tensor products. (Indeed, this is the quantity the entanglement-assisted classical capacity theorem establishes is equal to the entanglement-assisted classical capacity.)

Lemma 8.41 (Adami–Cerf). *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels. It holds that*

$$\begin{aligned} &\max_{\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1)\big) \\ &\quad = \max_{\sigma_0 \in \mathrm{D}(\mathcal{X}_0)} \big(\mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0)\big) + \max_{\sigma_1 \in \mathrm{D}(\mathcal{X}_1)} \big(\mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1)\big). \end{aligned} \tag{8.222}$$

Proof. Choose isometries $A_0 \in \mathrm{U}(\mathcal{X}_0, \mathcal{Y}_0 \otimes \mathcal{Z}_0)$ and $A_1 \in \mathrm{U}(\mathcal{X}_1, \mathcal{Y}_1 \otimes \mathcal{Z}_1)$, for an appropriate choice of complex Euclidean spaces $\mathcal{Z}_0$ and $\mathcal{Z}_1$, so that Stinespring representations of Φ_0 and Φ_1 are obtained:

$$\Phi_0(X_0) = \mathrm{Tr}_{\mathcal{Z}_0}\bigl(A_0 X_0 A_0^*\bigr) \quad \text{and} \quad \Phi_1(X_1) = \mathrm{Tr}_{\mathcal{Z}_1}\bigl(A_1 X_1 A_1^*\bigr) \tag{8.223}$$

for all $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$. The channels $\Psi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Z}_0)$ and $\Psi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Z}_1)$ defined as

$$\Psi_0(X_0) = \mathrm{Tr}_{\mathcal{Y}_0}\bigl(A_0 X_0 A_0^*\bigr) \quad \text{and} \quad \Psi_1(X_1) = \mathrm{Tr}_{\mathcal{Y}_1}\bigl(A_1 X_1 A_1^*\bigr) \tag{8.224}$$

for all $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$ are therefore complementary to Φ_0 and Φ_1, respectively.

Now, consider registers X_0, X_1, Y_0, Y_1, Z_0, and Z_1 corresponding to the spaces $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, $\mathcal{Y}_1$, $\mathcal{Z}_0$, and $\mathcal{Z}_1$, respectively. Let $\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)$ be an arbitrary density operator. With respect to the state

$$(A_0 \otimes A_1)\sigma(A_0 \otimes A_1)^* \in \mathrm{D}(\mathcal{Y}_0 \otimes \mathcal{Z}_0 \otimes \mathcal{Y}_1 \otimes \mathcal{Z}_1) \tag{8.225}$$

of $(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1)$, one has that

$$\begin{aligned} &\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1) \\ &\quad = \mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1) + \mathrm{H}(\mathsf{Y}_0, \mathsf{Y}_1) - \mathrm{H}(\mathsf{Z}_0, \mathsf{Z}_1). \end{aligned} \tag{8.226}$$

For every state of $(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1)$, including the state (8.225), it holds that

$$\begin{aligned} &\mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1) - \mathrm{H}(\mathsf{Z}_0, \mathsf{Z}_1) \\ &\quad \leq \mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0) - \mathrm{H}(\mathsf{Z}_0) + \mathrm{H}(\mathsf{Y}_1, \mathsf{Z}_1) - \mathrm{H}(\mathsf{Z}_1); \end{aligned} \tag{8.227}$$

two applications of the strong subadditivity of the von Neumann entropy (Theorem 5.39) yield this inequality. It follows from the subadditivity of the von Neumann entropy (Theorem 5.26) that

$$\mathrm{H}(\mathsf{Y}_0, \mathsf{Y}_1) \leq \mathrm{H}(\mathsf{Y}_0) + \mathrm{H}(\mathsf{Y}_1). \tag{8.228}$$

Consequently,

$$\begin{aligned} &\mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0, \mathsf{Y}_1, \mathsf{Z}_1) + \mathrm{H}(\mathsf{Y}_0, \mathsf{Y}_1) - \mathrm{H}(\mathsf{Z}_0, \mathsf{Z}_1) \\ &\qquad \leq \bigl(\mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0) + \mathrm{H}(\mathsf{Y}_0) - \mathrm{H}(\mathsf{Z}_0)\bigr) \\ &\qquad\quad + \bigl(\mathrm{H}(\mathsf{Y}_1, \mathsf{Z}_1) + \mathrm{H}(\mathsf{Y}_1) - \mathrm{H}(\mathsf{Z}_1)\bigr). \end{aligned} \tag{8.229}$$

For $\sigma_0 = \sigma[\mathsf{X}_0]$ and $\sigma_1 = \sigma[\mathsf{X}_1]$, one has the equations

$$\begin{aligned} \mathrm{H}(\mathsf{Y}_0, \mathsf{Z}_0) + \mathrm{H}(\mathsf{Y}_0) - \mathrm{H}(\mathsf{Z}_0) &= \mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0), \\ \mathrm{H}(\mathsf{Y}_1, \mathsf{Z}_1) + \mathrm{H}(\mathsf{Y}_1) - \mathrm{H}(\mathsf{Z}_1) &= \mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1). \end{aligned} \tag{8.230}$$

It follows that

$$\begin{aligned} &\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1) \\ &\quad \leq \big(\mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0)\big) + \big(\mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1)\big). \end{aligned} \tag{8.231}$$

Maximizing over all $\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)$, one obtains the inequality

$$\begin{aligned} &\max_{\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1)\big) \\ &\qquad \leq \max_{\sigma_0 \in \mathrm{D}(\mathcal{X}_0)} \big(\mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0)\big) \\ &\qquad\qquad + \max_{\sigma_1 \in \mathrm{D}(\mathcal{X}_1)} \big(\mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1)\big). \end{aligned} \tag{8.232}$$

For the reverse inequality, it suffices to observe that

$$\begin{aligned} &\mathrm{H}(\sigma_0 \otimes \sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_0 \otimes \sigma_1; \Phi_0 \otimes \Phi_1) \\ &\quad = \mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0) + \mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1) \end{aligned} \tag{8.233}$$

for every choice of $\sigma_0 \in \mathrm{D}(\mathcal{X}_0)$ and $\sigma_1 \in \mathrm{D}(\mathcal{X}_1)$, and therefore

$$\begin{aligned} &\max_{\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1)\big) \\ &\quad \geq \max_{\sigma_0 \in \mathrm{D}(\mathcal{X}_0)} \big(\mathrm{H}(\sigma_0) + \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0)\big) + \max_{\sigma_1 \in \mathrm{D}(\mathcal{X}_1)} \big(\mathrm{H}(\sigma_1) + \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1)\big), \end{aligned} \tag{8.234}$$

which completes the proof. □

A lower-bound on the Holevo capacity for flat states by dense coding

Next in the sequence of lemmas needed to prove the entanglement-assisted classical capacity theorem is the following lemma, which establishes a lower bound on the entanglement-assisted Holevo capacity of a given channel. Its proof may be viewed an application of dense coding (q.v. Section 6.3.1).

Lemma 8.42. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let $\Pi \in \mathrm{Proj}(\mathcal{X})$ be a nonzero projection operator, and let $\omega = \Pi / \operatorname{Tr}(\Pi)$. It holds that*

$$\chi_{\mathrm{E}}(\Phi) \geq \mathrm{H}(\omega) + \mathrm{I}_{\mathrm{c}}(\omega; \Phi). \tag{8.235}$$

Proof. Let $m = \operatorname{rank}(\Pi)$, let $\mathcal{W} = \mathbb{C}^{\mathbb{Z}_m}$, let $V \in \mathrm{U}(\mathcal{W}, \mathcal{X})$ be any isometry satisfying $VV^* = \Pi$, and let

$$\tau = \frac{1}{m} \operatorname{vec}(V) \operatorname{vec}(V)^* \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W}). \tag{8.236}$$

Recall the collection of discrete Weyl operators

$$\{W_{a,b} : a, b \in \mathbb{Z}_m\} \subset \mathrm{U}(\mathcal{W}), \tag{8.237}$$

as defined in Section 4.1.2 of Chapter 4, and define a collection of unitary channels

$$\{\Psi_{a,b} : a, b \in \mathbb{Z}_m\} \subseteq \mathrm{C}(\mathcal{W}) \tag{8.238}$$

in correspondence with these operators:

$$\Psi_{a,b}(Y) = W_{a,b} Y W_{a,b}^* \tag{8.239}$$

for each $Y \in \mathrm{L}(\mathcal{W})$. Finally, consider the ensemble

$$\eta : \mathbb{Z}_m \times \mathbb{Z}_m \to \mathrm{Pos}(\mathcal{X} \otimes \mathcal{W}) \tag{8.240}$$

defined as

$$\eta(a, b) = \frac{1}{m^2} \left(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Psi_{a,b}\right)(\tau), \tag{8.241}$$

for all $(a, b) \in \mathbb{Z}_m \times \mathbb{Z}_m$.

It holds that

$$\begin{aligned} &\mathrm{H}\left(\frac{1}{m^2} \sum_{a,b \in \mathbb{Z}_m} (\Phi \otimes \Psi_{a,b})(\tau)\right) \\ &\qquad = \mathrm{H}\left(\Phi(\omega) \otimes \frac{\mathbb{1}_{\mathcal{W}}}{m}\right) = \mathrm{H}\big(\Phi(\omega)\big) + \mathrm{H}(\omega) \end{aligned} \tag{8.242}$$

and

$$\begin{aligned} \frac{1}{m^2} \sum_{a,b \in \mathbb{Z}_n} \mathrm{H}\big((\Phi \otimes \Psi_{a,b})(\tau)\big) &= \mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\tau)\big) \\ &= \mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})})\big(\operatorname{vec}\big(\sqrt{\omega}\big) \operatorname{vec}\big(\sqrt{\omega}\big)^*\big)\big), \end{aligned} \tag{8.243}$$

from which it follows that

$$\chi\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\eta)\big) = \mathrm{H}(\omega) + \mathrm{I}_{\mathrm{C}}(\omega; \Phi). \tag{8.244}$$

Moreover, η is constant with respect to $\mathcal{W}$, as is evident from the fact that

$$\mathrm{Tr}_{\mathcal{X}}(\eta(a,b)) = \frac{1}{m^3}\mathbb{1}_{\mathcal{W}} \tag{8.245}$$

for each choice of $(a,b) \in \mathbb{Z}_m \times \mathbb{Z}_m$. It therefore holds that

$$\chi_{\mathrm{E}}(\Phi) \geq \chi\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\eta)\big) = \mathrm{H}(\omega) + \mathrm{I}_{\mathrm{C}}(\omega;\Phi), \tag{8.246}$$

which completes the proof. □

An upper-bound on the Holevo capacity

The final lemma needed for the proof of the entanglement-assisted classical capacity theorem establishes an upper bound on the entanglement-assisted Holevo capacity of a channel.

Lemma 8.43. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel. Also let $\mathcal{W}$ be a complex Euclidean space, let Σ be an alphabet, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X}\otimes\mathcal{W})$ be an ensemble that is constant with respect to $\mathcal{W}$, and let*

$$\sigma = \sum_{a\in\Sigma} \mathrm{Tr}_{\mathcal{W}}(\eta(a)). \tag{8.247}$$

It holds that

$$\chi\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\eta)\big) \leq \mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}(\sigma;\Phi). \tag{8.248}$$

Proof. Assume that $\mathcal{Z}$ is a complex Euclidean space and $A \in \mathrm{U}(\mathcal{X},\mathcal{Y}\otimes\mathcal{Z})$ is an isometry for which

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\big(AXA^*\big) \tag{8.249}$$

for all $X \in \mathrm{L}(\mathcal{X})$. The channel $\Psi \in \mathrm{C}(\mathcal{X},\mathcal{Z})$ defined by

$$\Psi(X) = \mathrm{Tr}_{\mathcal{Y}}\big(AXA^*\big) \tag{8.250}$$

for all $X \in \mathrm{L}(\mathcal{X})$ is therefore complementary to Φ. By Proposition 8.17, it follows that

$$\mathrm{I}_{\mathrm{C}}(\sigma;\Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)). \tag{8.251}$$

It therefore suffices to prove that

$$\chi\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\eta)\big) \leq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)). \tag{8.252}$$

By the assumption that η is constant with respect to $\mathcal{W}$, Proposition 8.12 implies that there must exist a complex Euclidean space $\mathcal{V}$, a collection of channels

$$\{\Xi_a : a \in \Sigma\} \subseteq \mathrm{C}(\mathcal{V}, \mathcal{X}), \tag{8.253}$$

a unit vector $u \in \mathcal{V} \otimes \mathcal{W}$, and a probability vector $p \in \mathcal{P}(\Sigma)$ such that

$$\eta(a) = \big(\Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(uu^*) \tag{8.254}$$

for every $a \in \Sigma$. Assume hereafter that such a choice for these objects has been fixed, and define states $\tau \in \mathrm{D}(\mathcal{W})$ and $\xi \in \mathrm{D}(\mathcal{V})$ as

$$\tau = \mathrm{Tr}_{\mathcal{V}}(uu^*) \quad \text{and} \quad \xi = \mathrm{Tr}_{\mathcal{W}}(uu^*). \tag{8.255}$$

It may be noted that

$$\sigma = \sum_{a \in \Sigma} p(a) \Xi_a(\xi). \tag{8.256}$$

Let $\mathcal{U}$ be a complex Euclidean space such that $\dim(\mathcal{U}) = \dim(\mathcal{V} \otimes \mathcal{X})$, and select a collection of isometries

$$\{B_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{V}, \mathcal{X} \otimes \mathcal{U}) \tag{8.257}$$

satisfying

$$\Xi_a(V) = \mathrm{Tr}_{\mathcal{U}}(B_a V B_a^*) \tag{8.258}$$

for every $V \in \mathrm{L}(\mathcal{V})$.

Assume momentarily that $a \in \Sigma$ has been fixed, and define a unit vector

$$v_b = (A \otimes \mathbb{1}_{\mathcal{U}} \otimes \mathbb{1}_{\mathcal{W}})(B_a \otimes \mathbb{1}_{\mathcal{W}})u \in \mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{U} \otimes \mathcal{W}. \tag{8.259}$$

Let Y, Z, U, and W be registers having corresponding complex Euclidean spaces $\mathcal{Y}$, $\mathcal{Z}$, $\mathcal{U}$, and $\mathcal{W}$, and consider the situation in which the compound register $(\mathsf{Y}, \mathsf{Z}, \mathsf{U}, \mathsf{W})$ is in the pure state $v_b v_b^*$. The following equalities may be verified:

$$\begin{aligned} \mathrm{H}(\mathsf{W}) &= \mathrm{H}(\tau), \\ \mathrm{H}(\mathsf{Y}, \mathsf{W}) &= \mathrm{H}\big(\big(\Phi \Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(uu^*)\big), \\ \mathrm{H}(\mathsf{U}, \mathsf{W}) &= \mathrm{H}(\mathsf{Y}, \mathsf{Z}) = \mathrm{H}\big(\Xi_a(\xi)\big), \\ \mathrm{H}(\mathsf{Y}, \mathsf{U}, \mathsf{W}) &= \mathrm{H}(\mathsf{Z}) = \mathrm{H}\big((\Psi \Xi_a)(\xi)\big). \end{aligned} \tag{8.260}$$

By the strong subadditivity of the von Neumann entropy (Theorem 5.39), it holds that

$$\mathrm{H}(\mathsf{W}) - \mathrm{H}(\mathsf{Y}, \mathsf{W}) \leq \mathrm{H}(\mathsf{U}, \mathsf{W}) - \mathrm{H}(\mathsf{Y}, \mathsf{U}, \mathsf{W}), \tag{8.261}$$

and therefore

$$\mathrm{H}(\tau) - \mathrm{H}\big((\Phi\Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(uu^*)\big) \leq \mathrm{H}\big(\Xi_a(\xi)\big) - \mathrm{H}\big((\Psi\Xi_a)(\xi)\big). \tag{8.262}$$

Finally, in accordance with the probability vector p, one may average the two sides of (8.262) over all $a \in \Sigma$, obtaining

$$\begin{aligned} \mathrm{H}(\tau) - \sum_{a\in\Sigma} p(a)\, \mathrm{H}\big((\Phi\Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(uu^*)\big) \\ \leq \sum_{a\in\Sigma} p(a)\big(\mathrm{H}\big(\Xi_a(\xi)\big) - \mathrm{H}\big((\Psi\Xi_a)(\xi)\big)\big). \end{aligned} \tag{8.263}$$

Lemma 8.40 therefore implies that

$$\mathrm{H}(\tau) - \sum_{a\in\Sigma} p(a)\, \mathrm{H}\big((\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho_a)\big) \leq \mathrm{H}(\sigma) - \mathrm{H}(\Psi(\sigma)). \tag{8.264}$$

By the subadditivity of the von Neumann entropy (Proposition 5.10) one has

$$\mathrm{H}\left(\sum_{a\in\Sigma} p(a)(\Phi\Xi_a \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(uu^*)\right) \leq \mathrm{H}(\Phi(\sigma)) + \mathrm{H}(\tau). \tag{8.265}$$

The inequality (8.252) follows from (8.264) and (8.265), which completes the proof. □

The entanglement-assisted classical capacity theorem

Finally, the entanglement-assisted classical capacity theorem will be stated, and proved through the use of the lemmas presented above.

Theorem 8.44 (Entanglement-assisted classical capacity theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that*

$$\mathrm{C}_{\mathrm{E}}(\Phi) = \max_{\sigma\in\mathrm{D}(\mathcal{X})} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}(\sigma\,;\,\Phi)\big). \tag{8.266}$$

Proof. By Lemmas 8.41 and 8.43, one may conclude that

$$\begin{aligned} \chi_{\mathrm{E}}(\Phi^{\otimes n}) &\leq \max_{\sigma\in\mathrm{D}(\mathcal{X}^{\otimes n})} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}\big(\sigma\,;\,\Phi^{\otimes n}\big)\big) \\ &= n \max_{\sigma\in\mathrm{D}(\mathcal{X})} \big(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}\big(\sigma\,;\,\Phi\big)\big) \end{aligned} \tag{8.267}$$

for every positive integer n. By Theorem 8.31, it therefore follows that

$$\mathrm{C}_{\mathrm{E}}(\Phi) = \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\bigl(\Phi^{\otimes n}\bigr)}{n} \leq \max_{\sigma\in\mathrm{D}(\mathcal{X})} \bigl(\mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}\bigl(\sigma\,;\,\Phi\bigr)\bigr). \tag{8.268}$$

For the reverse inequality, one may first choose a complex Euclidean space $\mathcal{Z}$ and an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y}\otimes\mathcal{Z})$ such that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}\bigl(AXA^*\bigr) \tag{8.269}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It holds that the channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$, defined by

$$\Psi(X) = \mathrm{Tr}_{\mathcal{Y}}\bigl(AXA^*\bigr) \tag{8.270}$$

for all $X \in \mathrm{L}(\mathcal{X})$, is complementary to Φ, so that Proposition 8.17 implies

$$\mathrm{I}_{\mathrm{C}}(\sigma\,;\,\Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) \tag{8.271}$$

for all $\sigma \in \mathrm{D}(\mathcal{X})$.

Let $\sigma \in \mathrm{D}(\mathcal{X})$ be any density operator, let $\delta > 0$ be chosen arbitrarily, and choose $\varepsilon > 0$ to be sufficiently small so that

$$(7\,\mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) + \mathrm{H}(\Psi(\sigma)))\varepsilon < \delta. \tag{8.272}$$

By Lemma 8.39, one may conclude that the inequality

$$\begin{aligned} \frac{1}{n}\Bigl(\mathrm{H}(\omega_{n,\varepsilon}) + \mathrm{H}(\Phi^{\otimes n}(\omega_{n,\varepsilon})) - \mathrm{H}(\Psi^{\otimes n}(\omega_{n,\varepsilon}))\Bigr) \\ \geq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) - \delta \end{aligned} \tag{8.273}$$

holds for all but finitely many positive integers n, where

$$\omega_{n,\varepsilon} = \frac{\Lambda_{n,\varepsilon}}{\mathrm{Tr}(\Lambda_{n,\varepsilon})} \tag{8.274}$$

and $\Lambda_{n,\varepsilon}$ denotes the ε-strongly typical projection with respect to any fixed spectral decomposition of σ. By Lemma 8.42, it follows that

$$\frac{\chi_{\mathrm{E}}(\Phi^{\otimes n})}{n} \geq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) - \delta \tag{8.275}$$

for all but finitely many positive integers n, and therefore

$$\mathrm{C}_{\mathrm{E}}(\Phi) = \lim_{n\to\infty} \frac{\chi_{\mathrm{E}}\bigl(\Phi^{\otimes n}\bigr)}{n} \geq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) - \delta. \tag{8.276}$$

As this inequality holds for all $\delta > 0$, one has

$$\mathrm{C}_{\mathrm{E}}(\Phi) \geq \mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)) = \mathrm{H}(\sigma) + \mathrm{I}_{\mathrm{C}}\bigl(\sigma\,;\,\Phi\bigr), \tag{8.277}$$

and maximizing over all $\sigma \in \mathrm{D}(\mathcal{X})$ completes the proof. □

8.2 Quantum information over quantum channels

This section is concerned with the capacity of quantum channels to transmit quantum information from a sender to a receiver. Along similar lines to the classical capacities considered in the previous section, one may consider the quantum capacity of a channel both when the sender and receiver share prior entanglement, used to assist the information transmission, and when they do not.

As it turns out, the capacity of a quantum channel to transmit quantum information with the assistance of entanglement is, in all cases, one-half of the entanglement-assisted classical capacity of the same channel. This fact is proved below through a combination of the teleportation and dense coding protocols discussed in Section 6.3.1. As the entanglement-assisted classical capacity has already been characterized by Theorem 8.44, a characterization of the capacity of a quantum channel to transmit quantum information with the assistance of entanglement follows directly. For this reason, the primary focus of the section is on an analysis of the capacity of quantum channels to transmit quantum information without the assistance of entanglement.

The first subsection below presents a definition of the quantum capacity of a channel, together with the closely related notion of a channel's capacity to generate shared entanglement. The second subsection presents a proof of the quantum capacity theorem, which characterizes the capacity of a given channel to transmit quantum information.

8.2.1 Definitions of quantum capacity and related notions

Definitions of the *quantum capacity* and the *entanglement-generation capacity* of a quantum channel are presented below, and it is proved that the two quantities coincide. The *entanglement-assisted quantum capacity* of a quantum channel is also defined, and its fairly straightforward relationship to the entanglement-assisted classical capacity of a channel is clarified.

The quantum capacity of a channel

Informally speaking, the quantum capacity of a channel is the number of qubits, on average, that can be accurately transmitted with each use of that channel. Like the capacities discussed in the previous section, the quantum capacity of a channel is defined in information-theoretic terms, referring to

a situation in which an asymptotically large number of channel uses, acting on a collection of possibly entangled registers, is made available.

The definition of quantum capacity that follows makes use of the same notions of an emulation of one channel by another (Definition 8.1) and of an ε-approximation of one channel by another (Definition 8.2) that were used in the previous section.

Definition 8.45 (Quantum capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Also let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for quantum information transmission through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$.
2. The *quantum capacity* of Φ, which is denoted $\mathrm{Q}(\Phi)$, is defined as the supremum of all achievable rates for quantum information transmission through Φ.

Similar to the classical capacities considered in the previous section, the argument through which Proposition 8.4 was proved yields an analogous proposition for the quantum capacity.

Proposition 8.46. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let k be a positive integer. It holds that $\mathrm{Q}(\Phi^{\otimes k}) = k\,\mathrm{Q}(\Phi)$.*

The entanglement generation capacity of a channel

The *entanglement generation capacity* of a channel is defined in a similar way to the quantum capacity, except that the associated task is more narrowly focused—by means of multiple, independent uses of a channel, a sender and receiver aim to establish a state, shared between them, having high fidelity with a maximally entangled state.

Definition 8.47 (Entanglement generation capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Also let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for entanglement generation through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, there exists a state $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ and a channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ such that

$$\mathrm{F}\Big(2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \big(\Xi \Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho)\Big) \geq 1 - \varepsilon. \tag{8.278}$$

2. The *entanglement generation capacity* of Φ, denoted $\mathrm{Q}_{\mathrm{EG}}(\Phi)$, is defined as the supremum of all achievable rates for entanglement generation through Φ.

Remark 8.48. For any choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, a unit vector $y \in \mathcal{Y}$, and a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, the maximum value for the fidelity $\mathrm{F}(yy^*, \Psi(\rho))$ over $\rho \in \mathrm{D}(\mathcal{X})$ is achieved when ρ is a pure state. It follows from this observation that the quantity $\mathrm{Q}_{\mathrm{EG}}(\Phi)$ would not change if the states $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ considered in the specification of achievable rates in Definition 8.47 are constrained to be pure states.

Equivalence of quantum capacity and entanglement generation capacity

The task associated with entanglement generation capacity would seem to be more specialized than the one associated with quantum capacity; it is apparent that the emulation of a close approximation to an identity channel allows a sender and receiver to generate a shared state having high fidelity with a maximally entangled state, but it is not clear that the reverse should be true. The relationship between entanglement generation and identity channel emulation provided by the following theorem allows one to prove that the reverse implication does indeed hold: the quantum capacity and entanglement generation capacity of any given channel always coincide.

Theorem 8.49. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $n = \dim(\mathcal{Y})$, and assume $\dim(\mathcal{Z}) \leq n/2$. Let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, let $u \in \mathcal{X} \otimes \mathcal{Y}$ be a unit vector, let $\varepsilon \geq 0$ be a nonnegative real number, and assume the inequality*

$$\mathrm{F}\left(\frac{1}{n} \operatorname{vec}(\mathbb{1}_{\mathcal{Y}}) \operatorname{vec}(\mathbb{1}_{\mathcal{Y}})^*, \big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(uu^*)\right) \geq 1 - \varepsilon \tag{8.279}$$

is satisfied. The channel Φ emulates a δ-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}$ for $\delta = \big(4\sqrt{2}\big)\varepsilon^{\frac{1}{4}}$.

Proof. Let $A \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ be the operator defined by the equation $\mathrm{vec}(A) = u$, and let

$$A = \sum_{k=1}^{r} \sqrt{p_k}\, x_k y_k^* \tag{8.280}$$

be a singular value decomposition of A, where $r \leq n$ is the rank of A, $\{x_1, \ldots, x_r\} \subset \mathcal{X}$ and $\{y_1, \ldots, y_r\} \subset \mathcal{Y}$ are orthonormal sets, and $(p_1, \ldots, p_r)$ is a probability vector. Define $W \in \mathrm{L}(\mathcal{Y}, \mathcal{X})$ as

$$W = \sum_{k=1}^{r} x_k y_k^*, \tag{8.281}$$

and define a unit vector $v \in \mathcal{X} \otimes \mathcal{Y}$ as

$$v = \frac{1}{\sqrt{r}} \mathrm{vec}(W). \tag{8.282}$$

By the monotonicity of the fidelity function under partial tracing, one has

$$\begin{aligned} \mathrm{F}(uu^*, vv^*) &= \frac{1}{\sqrt{r}} \sum_{k=1}^{r} \sqrt{p_k} \geq \frac{1}{\sqrt{n}} \sum_{k=1}^{r} \sqrt{p_k} = \mathrm{F}\left(\frac{1}{n} \mathbb{1}_{\mathcal{Y}}, \mathrm{Tr}_{\mathcal{X}}(uu^*)\right) \\ &\geq \mathrm{F}\left(\frac{1}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{Y}}) \mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*, \left(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(uu^*)\right) \geq 1 - \varepsilon. \end{aligned} \tag{8.283}$$

Consequently, by Theorems 3.30 and 3.32, one has

$$\begin{aligned} &\mathrm{F}\left(\frac{1}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{Y}}) \mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*, \left(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(vv^*)\right) + 1 \\ &\quad \geq \mathrm{F}\left(\frac{1}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{Y}}) \mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*, \left(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(uu^*)\right)^2 + \mathrm{F}\left(vv^*, uu^*\right)^2 \\ &\quad \geq 2(1 - \varepsilon)^2, \end{aligned} \tag{8.284}$$

and therefore

$$\mathrm{F}\left(\frac{1}{n} \mathrm{vec}(\mathbb{1}_{\mathcal{Y}}) \mathrm{vec}(\mathbb{1}_{\mathcal{Y}})^*, \left(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(vv^*)\right) \geq 1 - 4\varepsilon. \tag{8.285}$$

Next, define a projection operator $\Pi_r = W^* W \in \mathrm{Proj}(\mathcal{Y})$ and define $\mathcal{V}_r = \mathrm{im}(\Pi_r)$. For each choice of k beginning with r and decreasing to 1, choose $w_k \in \mathcal{V}_k$ to be a unit vector that minimizes the quantity

$$\alpha_k = \left\langle w_k w_k^*, \Phi\left(W w_k w_k^* W^*\right)\right\rangle, \tag{8.286}$$

and define

$$\mathcal{V}_{k-1} = \{z \in \mathcal{V}_k \,:\, \langle w_k, z\rangle = 0\}. \tag{8.287}$$

Observe that $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r$ and that $\{w_1, \ldots, w_k\}$ is an orthonormal basis for $\mathcal{V}_k$, for each $k \in \{1, \ldots, r\}$. In particular, it holds that

$$v = \frac{1}{\sqrt{r}}(W \otimes \mathbb{1}_{\mathcal{Y}})\operatorname{vec}(\Pi_r) = \frac{1}{\sqrt{r}}\sum_{k=1}^{r} W w_k \otimes \overline{w_k}. \tag{8.288}$$

At this point, a calculation reveals that

$$\begin{aligned} &\mathrm{F}\left(\frac{1}{n}\operatorname{vec}(\mathbb{1}_{\mathcal{Y}})\operatorname{vec}(\mathbb{1}_{\mathcal{Y}})^*, \left(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(vv^*)\right)^2 \\ &\qquad = \frac{1}{nr}\sum_{j,k\in\{1,\ldots,r\}} \left\langle w_j w_k^*, \Phi(W w_j w_k^* W^*)\right\rangle. \end{aligned} \tag{8.289}$$

By the complete positivity of Φ, one may conclude that

$$\begin{aligned} &\left|\left\langle w_j w_k^*, \Phi(W w_j w_k^* W^*)\right\rangle\right| \\ &\qquad \leq \sqrt{\left\langle w_j w_j^*, \Phi(W w_j w_j^* W^*)\right\rangle}\sqrt{\left\langle w_k w_k^*, \Phi(W w_k w_k^* W^*)\right\rangle} \\ &\qquad = \sqrt{\alpha_j \alpha_k}, \end{aligned} \tag{8.290}$$

for each choice of $j, k \in \{1, \ldots, r\}$. Therefore, by the triangle inequality one may conclude that

$$\mathrm{F}\left(\frac{1}{n}\operatorname{vec}(\mathbb{1}_{\mathcal{Y}})\operatorname{vec}(\mathbb{1}_{\mathcal{Y}})^*, \left(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\right)(vv^*)\right) \leq \frac{1}{\sqrt{nr}}\sum_{k=1}^{r}\sqrt{\alpha_k}. \tag{8.291}$$

Applying the Cauchy–Schwarz inequality, one obtains

$$\frac{1}{\sqrt{nr}}\sum_{k=1}^{r}\sqrt{\alpha_k} \leq \sqrt{\frac{1}{n}\sum_{k=1}^{r}\alpha_k}, \tag{8.292}$$

and therefore

$$\frac{1}{n}\sum_{k=1}^{r}\alpha_k \geq (1-4\varepsilon)^2 \geq 1 - 8\varepsilon. \tag{8.293}$$

Now choose $m \in \{0, \ldots, r\}$ to be the maximum value of m for which it holds that $\alpha_k \geq 1 - 16\varepsilon$ for all $k \leq m$. By (8.293) it necessarily holds that $m \geq n/2$. By the definition of the values $\alpha_1, \ldots, \alpha_r$, one may conclude that

$$\left\langle ww^*, \Phi(Www^*W^*)\right\rangle \geq 1 - 16\varepsilon \tag{8.294}$$

for every unit vector $w \in \mathcal{V}_m$. Finally, let $V \in \mathrm{U}(\mathcal{Z}, \mathcal{Y})$ be any isometry with the property that $\mathrm{im}(V) \subseteq \mathcal{V}_m$. Such an isometry exists by the assumption that $\dim(\mathcal{Z}) \leq n/2$ together with the fact that $n/2 \leq m = \dim(\mathcal{V}_m)$. Let $\Xi_{\mathrm{E}} \in \mathrm{C}(\mathcal{Z}, \mathcal{X})$ and $\Xi_{\mathrm{D}} \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ be channels of the form

$$\Xi_{\mathrm{E}}(Z) = WVZV^*W^* + \Psi_{\mathrm{E}}(Z) \quad \text{and} \quad \Xi_{\mathrm{D}}(Y) = V^*YV + \Psi_{\mathrm{D}}(Y), \tag{8.295}$$

for all $Z \in \mathrm{L}(\mathcal{Z})$ and $Y \in \mathrm{L}(\mathcal{Y})$, where $\Psi_{\mathrm{E}} \in \mathrm{CP}(\mathcal{Z}, \mathcal{X})$ and $\Psi_{\mathrm{D}} \in \mathrm{CP}(\mathcal{Y}, \mathcal{Z})$ are completely positive maps that cause Ξ_{E} and Ξ_{D} to be trace-preserving. For every unit vector $z \in \mathcal{Z}$ it holds that

$$\left\langle zz^*, (\Xi_{\mathrm{D}}\Phi\Xi_{\mathrm{E}})(zz^*) \right\rangle \geq \left\langle Vzz^*V^*, \Phi(WVzz^*V^*W^*) \right\rangle \geq 1 - 16\varepsilon, \tag{8.296}$$

and therefore one has

$$\left\| zz^* - (\Xi_{\mathrm{D}}\Phi\Xi_{\mathrm{E}})(zz^*) \right\|_1 \leq 8\sqrt{\varepsilon} \tag{8.297}$$

by one of the Fuchs–van de Graaf inequalities (Theorem 3.36). Applying Theorem 3.58, one therefore finds that

$$\left\| \Xi_{\mathrm{D}}\Phi\Xi_{\mathrm{E}} - \mathbb{1}_{\mathrm{L}(\mathcal{Z})} \right\|_1 \leq \left(4\sqrt{2}\right)\varepsilon^{\frac{1}{4}}, \tag{8.298}$$

which completes the proof. □

Theorem 8.50. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The quantum capacity and entanglement generation capacity of Φ are equal: $\mathrm{Q}(\Phi) = \mathrm{Q}_{\mathrm{E}}(\Phi)$.*

Proof. It will first be proved that $\mathrm{Q}(\Phi) \leq \mathrm{Q}_{\mathrm{E}}(\Phi)$, which is straightforward. If the quantum capacity of Φ is zero, there is nothing to prove, so it will be assumed that $\mathrm{Q}(\Phi) > 0$. Let $\alpha > 0$ be an achievable rate for quantum information transmission through Φ, and let $\varepsilon > 0$ be chosen arbitrarily.

Setting $\Gamma = \{0, 1\}$ and $\mathcal{Z} = \mathbb{C}^{\Gamma}$, one therefore has that the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$ for all but finitely many positive integers n and for all positive integers $m \leq \alpha n$. That is, for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, there must exist channels $\Xi_{\mathrm{E}} \in \mathrm{C}(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n})$ and $\Xi_{\mathrm{D}} \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ such that

$$\left\| \Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} - \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m} \right\|_1 < \varepsilon. \tag{8.299}$$

Supposing that n and m are positive integers for which such channels exist, one may consider the density operators

$$\tau = 2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^* \quad \text{and} \quad \rho = \bigl(\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\bigr)(\tau), \tag{8.300}$$

along with the channel $\Xi = \Xi_{\mathrm{D}}$. One of the Fuchs–van de Graaf inequalities (Theorem 3.36) implies that

$$\begin{aligned} \mathrm{F}\Bigl(\tau, \bigl(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\bigr)(\rho)\Bigr) &= \mathrm{F}\Bigl(\tau, \bigl(\Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\bigr)(\tau)\Bigr) \\ &\geq 1 - \frac{1}{2}\bigl\| \bigl(\Xi_{\mathrm{D}}\Phi^{\otimes n}\Xi_{\mathrm{E}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\bigr)(\tau) - \tau \bigr\|_1 \geq 1 - \frac{\varepsilon}{2}. \end{aligned} \tag{8.301}$$

As this is so for all but finitely many positive integers n and all positive integers $m \leq \alpha n$, it is the case that α is an achievable rate for entanglement generation through Φ. Taking the supremum over all achievable rates α for quantum communication through Φ, one obtains $\mathrm{Q}(\Phi) \leq \mathrm{Q}_{\mathrm{E}}(\Phi)$.

It remains to prove that $\mathrm{Q}_{\mathrm{E}}(\Phi) \leq \mathrm{Q}(\Phi)$. As for the reverse inequality just proved, there is nothing to prove if $\mathrm{Q}_{\mathrm{E}}(\Phi) = 0$, so it will be assumed that $\mathrm{Q}_{\mathrm{E}}(\Phi) > 0$. Let $\alpha > 0$ be an achievable rate for entanglement generation through Φ and let $\beta \in (0, \alpha)$ be chosen arbitrarily. It will be proved that β is an achievable rate for quantum communication through Φ. The required relation $\mathrm{Q}_{\mathrm{E}}(\Phi) \leq \mathrm{Q}(\Phi)$ follows by taking the supremum over all achievable rates α for entanglement generation through Φ and over all $\beta \in (0, \alpha)$.

Let $\varepsilon > 0$ be chosen arbitrarily and let $\delta = \varepsilon^4/1024$, so that $\bigl(4\sqrt{2}\bigr)\delta^{\frac{1}{4}} = \varepsilon$. One has that, for all but finitely many positive integers n and all positive integers $m \leq \alpha n$, that there exists a state $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ and a channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ such that

$$\mathrm{F}\bigl(2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \bigl(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\bigr)(\rho)\bigr) \geq 1 - \delta. \tag{8.302}$$

Fix n and $m \leq \alpha n$ to be positive integers for which this statement holds, and observe that the function

$$\begin{aligned} \rho \mapsto \mathrm{F}\bigl(2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \bigl(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\bigr)(\rho)\bigr)^2 \\ = \bigl\langle 2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \bigl(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\bigr)(\rho)\bigr\rangle \end{aligned} \tag{8.303}$$

must achieve its maximum value (over all density operators) on a pure state. Thus, there must exist a unit vector $u \in \mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m}$ so that the inequality (8.302) holds when $\rho = uu^*$. By Theorem 8.49, it follows that $\Phi^{\otimes n}$ emulates

an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes k}$ for every positive integer $k \leq m - 1$.

Under the assumption $n \geq 1/(\alpha - \beta)$, one has that $\beta n \leq \alpha n - 1$. Thus, for all but finitely many positive integers n and all positive integers $k \leq \beta n$, it holds that $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes k}$. As $\varepsilon > 0$ has been chosen arbitrarily, it follows that β is an achievable rate for quantum communication through Φ, which completes the proof. □

The entanglement-assisted quantum capacity of a channel

The entanglement-assisted quantum capacity of a channel, which will be proved is equal to one-half of its entanglement-assisted classical capacity, may be formally defined as follows.

Definition 8.51 (Entanglement-assisited quantum capacity of a channel)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Also let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$.

1. A value $\alpha \geq 0$ is an *achievable rate* for entanglement-assisted quantum information transmission through Φ if and only if (i) $\alpha = 0$, or (ii) $\alpha > 0$ and the following holds for every choice of a positive real number $\varepsilon > 0$: for all but finitely many positive integers n, and for all positive integers $m \leq \alpha n$, the channel $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$ with the assistance of entanglement.
2. The *entanglement-assisted quantum capacity* of Φ, denoted $\mathrm{Q}_{\mathrm{E}}(\Phi)$, is the supremum of all achievable rates for entanglement-assisted quantum information transmission through Φ.

Proposition 8.52. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It holds that*

$$\mathrm{Q}_{\mathrm{E}}(\Phi) = \frac{1}{2}\mathrm{C}_{\mathrm{E}}(\Phi). \tag{8.304}$$

Proof. Assume α is an achievable rate for entanglement-assisted classical communication through Φ. If $\alpha = 0$, then $\alpha/2$ is trivially an achievable rate for entanglement-assisted quantum information transmission through Φ. It will be observed that if $\alpha > 0$, then $\alpha/2 - \delta$ is an achievable rate for entanglement-assisted quantum communication through Φ for all real numbers $\delta \in (0, \alpha/2)$. Taking the supremum over all achievable rates α for

entanglement-assisted classical communication through Φ and the infimum over all $\delta \in (0, \alpha/2)$, one obtains

$$\mathrm{Q}_{\mathrm{E}}(\Phi) \geq \frac{1}{2}\mathrm{C}_{\mathrm{E}}(\Phi). \tag{8.305}$$

Suppose n and $m \leq \alpha n$ are positive integers and $\varepsilon > 0$ is a positive real number such that $\Phi^{\otimes n}$ emulates an ε-approximation to the channel $\Delta^{\otimes m}$, where $\Delta \in \mathrm{C}(\mathcal{Z})$ denotes the completely dephasing channel as usual. Let $k = \lfloor m/2 \rfloor$, and consider the maximally entangled state

$$\tau = 2^{-k}\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes k}\big)\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes k}\big)^{*}. \tag{8.306}$$

By tensoring τ with the state ξ used for the emulation of an ε-approximation to $\Delta^{\otimes m}$ by $\Phi^{\otimes n}$, one may define a new channel $\Psi \in \mathrm{C}(\mathcal{Z}^{\otimes k})$ through the use of the traditional teleportation protocol (q.v. Example 6.56 in Section 6.3.1), but where the classical communication channel required for teleportation is replaced by the ε-approximation to the channel $\Delta^{\otimes m}$ emulated by $\Phi^{\otimes n}$. It holds that Ψ is an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes k}$.

One therefore has that, for all $\varepsilon > 0$, for all but finitely many values of n, and for all $k \leq (\alpha n - 1)/2$, that $\Phi^{\otimes n}$ emulates an ε-approximation to the identity channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes k}$ through the assistance of entanglement. For every $\delta \in (0, \alpha/2)$, it is therefore the case that $\alpha/2 - \delta$ is an achievable rate for entanglement-assisted quantum communication through Φ, as required.

Now assume α is an achievable rate for entanglement-assisted quantum communication through Φ. It will be proved that 2α is an achievable rate for entanglement-assisted classical communication through Φ. This statement is trivial in the case $\alpha = 0$, so it will be assumed that $\alpha > 0$. The proof is essentially the same as the reverse direction just considered, with dense coding replacing teleportation.

Suppose that n and $m \leq \alpha n$ are positive integers and $\varepsilon > 0$ is a positive real number such that $\Phi^{\otimes n}$ emulates an ε-approximation to $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$. Using the maximally entangled state

$$\tau = 2^{-m}\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)^{*}, \tag{8.307}$$

tensored with the state ξ used for the emulation of $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$ by $\Phi^{\otimes n}$, one may define a new channel $\Psi \in \mathrm{C}(\mathcal{Z}^{\otimes 2m})$ through the traditional dense coding protocol (q.v. Example 6.61 in Section 6.3.1), where the quantum channel

required for dense coding is replaced by the ε-approximation to the channel $\mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}$ emulated by $\Phi^{\otimes n}$. It holds that Ψ is an ε-approximation to $\Delta^{\otimes 2m}$.

It therefore holds that, for all $\varepsilon > 0$, for all but finitely many values of n, and for all $m \leq \alpha n$, that $\Phi^{\otimes n}$ emulates an ε-approximation to the channel $\Delta^{\otimes 2m}$, which implies that 2α is an achievable rate for entanglement-assisted classical communication through Φ. The inequality

$$\mathrm{C}_{\mathrm{E}}(\Phi) \geq 2\,\mathrm{Q}_{\mathrm{E}}(\Phi) \tag{8.308}$$

is obtained when one takes the supremum over all achievable rates α for entanglement-assisted quantum communication through Φ.

The equality (8.304) therefore holds, which completes the proof. □

8.2.2 The quantum capacity theorem

The purpose of the present subsection is to state and prove the quantum capacity theorem, which yields an expression for the quantum capacity of a given channel. Similar to the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30), the expression that is obtained from the quantum capacity theorem includes a regularization over an increasing number of uses of a given channel.

The subsections that follow include statements and proofs of various lemmas that will be used to prove the quantum capacity theorem, along with the statement and proof of the theorem itself.

A decoupling lemma

The first of several lemmas that will be used to prove the quantum capacity theorem concerns a phenomenon known as *decoupling*. Informally speaking, this is the phenomenon whereby the action of a sufficiently noisy channel on a randomly chosen subspace of its input space can be expected not only to destroy entanglement with an auxiliary system, but to destroy classical correlations as well. The lemma that follows proves a fact along these lines that is specialized to the task at hand.

Lemma 8.53. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{W}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $n = \dim(\mathcal{X})$ and $m = \dim(\mathcal{Z})$, and assume $m \leq n \leq \dim(\mathcal{Y} \otimes \mathcal{W})$. Assume moreover that $V \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$ and $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ are isometries, and define density operators*

$\xi \in \mathrm{D}(\mathcal{W} \otimes \mathcal{X})$ and $\rho_U \in \mathrm{D}(\mathcal{W} \otimes \mathcal{Z})$, for each unitary operator $U \in \mathrm{U}(\mathcal{X})$, as follows:

$$\begin{aligned} \xi &= \frac{1}{n} \operatorname{Tr}_{\mathcal{Y}}\bigl(\operatorname{vec}(A)\operatorname{vec}(A)^*\bigr), \\ \rho_U &= \frac{1}{m} \operatorname{Tr}_{\mathcal{Y}}\bigl(\operatorname{vec}(AUV)\operatorname{vec}(AUV)^*\bigr). \end{aligned} \tag{8.309}$$

It holds that

$$\int \bigl\| \rho_U - \operatorname{Tr}_{\mathcal{Z}}(\rho_U) \otimes \omega \bigr\|_2^2 \, \mathrm{d}\eta(U) \leq \operatorname{Tr}(\xi^2), \tag{8.310}$$

for η denoting the Haar measure on $\mathrm{U}(\mathcal{X})$ and $\omega \in \mathrm{D}(\mathcal{Z})$ being the completely mixed state on $\mathcal{Z}$.

Proof. Observe first that

$$\bigl\| \rho_U - \operatorname{Tr}_{\mathcal{Z}}(\rho_U) \otimes \omega \bigr\|_2^2 = \operatorname{Tr}\bigl(\rho_U^2\bigr) - \frac{1}{m} \operatorname{Tr}\Bigl(\bigl(\operatorname{Tr}_{\mathcal{Z}}(\rho_U)\bigr)^2 \Bigr). \tag{8.311}$$

The lemma requires a bound on the integral of the expression represented by (8.311) over all U, and toward this goal the two terms on the right-hand side of that equation will be integrated separately.

To integrate the first term on the right-hand side of (8.311), let Γ be the alphabet for which $\mathcal{Y} = \mathbb{C}^{\Gamma}$, define $B_a = \bigl(e_a^* \otimes \mathbb{1}_{\mathcal{W}}\bigr) A$ for each $a \in \Gamma$, and observe that

$$\rho_U = \frac{1}{m} \sum_{a \in \Gamma} \operatorname{vec}(B_a U V) \operatorname{vec}(B_a U V)^*. \tag{8.312}$$

It therefore holds that

$$\begin{aligned} \operatorname{Tr}\bigl(\rho_U^2\bigr) &= \frac{1}{m^2} \sum_{a,b \in \Gamma} \bigl| \operatorname{Tr}\bigl(V^* U^* B_a^* B_b U V\bigr) \bigr|^2 \\ &= \frac{1}{m^2} \sum_{a,b \in \Gamma} \operatorname{Tr}\bigl(V^* U^* B_a^* B_b U V \otimes V^* U^* B_b^* B_a U V\bigr) \\ &= \Biggl\langle U V V^* U^* \otimes U V V^* U^*, \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \Biggr\rangle. \end{aligned} \tag{8.313}$$

Integrating over all $U \in \mathrm{U}(\mathcal{X})$ yields

$$\int \operatorname{Tr}\bigl(\rho_U^2\bigr) \, \mathrm{d}\eta(U) = \Biggl\langle \Xi(V V^* \otimes V V^*), \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \Biggr\rangle, \tag{8.314}$$

for $\Xi \in \mathrm{C}(\mathcal{X} \otimes \mathcal{X})$ denoting the Werner twirling channel (q.v. Example 7.26 in the previous chapter). Making use of the expression

$$\Xi(X) = \frac{2}{n(n+1)} \langle \Pi_{\mathcal{X} \circledcirc \mathcal{X}}, X \rangle \, \Pi_{\mathcal{X} \circledcirc \mathcal{X}} + \frac{2}{n(n-1)} \langle \Pi_{\mathcal{X} \circledcirc \mathcal{X}}, X \rangle \, \Pi_{\mathcal{X} \circledcirc \mathcal{X}}, \tag{8.315}$$

which holds for every $X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{X})$, and observing the equations

$$\big\langle \Pi_{\mathcal{X} \circledcirc \mathcal{X}}, VV^* \otimes VV^* \big\rangle = \frac{m(m+1)}{2}, \tag{8.316}$$

$$\big\langle \Pi_{\mathcal{X} \circledcirc \mathcal{X}}, VV^* \otimes VV^* \big\rangle = \frac{m(m-1)}{2}, \tag{8.317}$$

it follows that

$$\begin{aligned} &\int \mathrm{Tr}(\rho_U^2) \, \mathrm{d}\eta(U) \\ &\quad = \left\langle \frac{m(m+1)}{n(n+1)} \Pi_{\mathcal{X} \circledcirc \mathcal{X}} + \frac{m(m-1)}{n(n-1)} \Pi_{\mathcal{X} \circledcirc \mathcal{X}}, \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \right\rangle. \end{aligned} \tag{8.318}$$

A similar methodology can be used to integrate the second term on the right-hand side of (8.311). In particular, one has

$$\mathrm{Tr}_{\mathcal{Z}}(\rho_U) = \frac{1}{m} \sum_{a \in \Gamma} B_a U V V^* U^* B_a^*, \tag{8.319}$$

and therefore

$$\begin{aligned} &\mathrm{Tr}\Big(\big(\mathrm{Tr}_{\mathcal{Z}}(\rho_U)\big)^2 \Big) \\ &\quad = \frac{1}{m^2} \sum_{a,b \in \Gamma} \mathrm{Tr}\big(V^* U^* B_a^* B_b U V V^* U^* B_b^* B_a U V \big) \\ &\quad = \left\langle W_{\mathcal{Z}}, \frac{1}{m^2} \sum_{a,b \in \Gamma} V^* U^* B_a^* B_b U V \otimes V^* U^* B_b^* B_a U V \right\rangle \\ &\quad = \left\langle (UV \otimes UV) W_{\mathcal{Z}} (UV \otimes UV)^*, \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \right\rangle, \end{aligned} \tag{8.320}$$

where $W_{\mathcal{Z}} \in \mathrm{U}(\mathcal{Z} \otimes \mathcal{Z})$ denotes the swap operator on $\mathcal{Z} \otimes \mathcal{Z}$, and the second equality has used the identity $\langle W_{\mathcal{Z}}, X \otimes Y \rangle = \mathrm{Tr}(XY)$. Integrating over all $U \in \mathrm{U}(\mathcal{X})$ yields

$$\begin{aligned} &\int \mathrm{Tr}\Big(\big(\mathrm{Tr}_{\mathcal{Z}}(\rho_U)\big)^2 \Big) \, \mathrm{d}\eta(U) \\ &\quad = \left\langle \Xi\big((V \otimes V) W_{\mathcal{Z}} (V \otimes V)^* \big), \frac{1}{m^2} \sum_{a,b \in \Gamma} B_a^* B_b \otimes B_b^* B_a \right\rangle. \end{aligned} \tag{8.321}$$

By making use of the equations

$$\begin{aligned}\left\langle \Pi_{\mathcal{X}\otimes\mathcal{X}}, (V\otimes V)W_{\mathcal{Z}}(V\otimes V)^*\right\rangle &= \frac{m(m+1)}{2},\\ \left\langle \Pi_{\mathcal{X}\otimes\mathcal{X}}, (V\otimes V)W_{\mathcal{Z}}(V\otimes V)^*\right\rangle &= -\frac{m(m-1)}{2},\end{aligned} \tag{8.322}$$

and performing a similar calculation to the one above, one finds that

$$\begin{aligned}&\int \operatorname{Tr}\Big(\big(\operatorname{Tr}_{\mathcal{Z}}(\rho_U)\big)^2\Big)\,\mathrm{d}\eta(U)\\ &\quad= \left\langle \frac{m(m+1)}{n(n+1)}\Pi_{\mathcal{X}\otimes\mathcal{X}} - \frac{m(m-1)}{n(n-1)}\Pi_{\mathcal{X}\otimes\mathcal{X}}, \frac{1}{m^2}\sum_{a,b\in\Gamma} B_a^*B_b\otimes B_b^*B_a\right\rangle.\end{aligned} \tag{8.323}$$

Combining (8.318) and (8.323), together with some algebra, it follows that

$$\begin{aligned}&\int \left\|\rho_U - \operatorname{Tr}_{\mathcal{Z}}(\rho_U)\otimes\omega\right\|_2^2 \mathrm{d}\eta(U)\\ &\qquad= \frac{m^2-1}{n^2-1}\left\langle \mathbb{1}_{\mathcal{X}}\otimes\mathbb{1}_{\mathcal{X}} - \frac{1}{n}W_{\mathcal{X}}, \frac{1}{m^2}\sum_{a,b\in\Gamma} B_a^*B_b\otimes B_b^*B_a\right\rangle,\end{aligned} \tag{8.324}$$

where $W_{\mathcal{X}}$ denotes the swap operator on $\mathcal{X}\otimes\mathcal{X}$. By similar calculations to (8.313) and (8.320) above, but replacing U and V by $\mathbb{1}_{\mathcal{X}}$, it may be verified that

$$\operatorname{Tr}(\xi^2) = \frac{1}{n^2}\operatorname{Tr}\left(\sum_{a,b\in\Gamma} B_a^*B_b\otimes B_b^*B_a\right) \tag{8.325}$$

and

$$\operatorname{Tr}\Big(\big(\operatorname{Tr}_{\mathcal{X}}(\xi)\big)^2\Big) = \frac{1}{n^2}\left\langle W_{\mathcal{X}}, \sum_{a,b\in\Gamma} B_a^*B_b\otimes B_b^*B_a\right\rangle. \tag{8.326}$$

Consequently,

$$\begin{aligned}&\int \left\|\rho_U - \operatorname{Tr}_{\mathcal{Z}}(\rho_U)\otimes\omega\right\|_2^2 \mathrm{d}\eta(U)\\ &\qquad= \frac{1-m^{-2}}{1-n^{-2}}\left(\operatorname{Tr}(\xi^2) - \frac{1}{n}\operatorname{Tr}\Big(\big(\operatorname{Tr}_{\mathcal{X}}(\xi)\big)^2\Big)\right) \leq \operatorname{Tr}(\xi^2),\end{aligned} \tag{8.327}$$

as required. □

A lower-bound on entanglement generation decoding fidelity

The next lemma is used, within the proof of the quantum capacity theorem, to infer the existence of a decoding channel for the task of entanglement generation. This inference is based on a calculation involving a Stinespring representation of the channel through which entanglement generation is to be considered.

Lemma 8.54. *Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{W}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $m = \dim(\mathcal{Z})$, and assume that $m \leq \dim(\mathcal{X}) \leq \dim(\mathcal{Y} \otimes \mathcal{W})$. Also let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ and $V \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$ be isometries, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be the channel defined by*

$$\Phi(X) = \mathrm{Tr}_{\mathcal{W}}\big(AXA^*\big) \tag{8.328}$$

for all $X \in \mathrm{L}(\mathcal{X})$, let $\rho \in \mathrm{D}(\mathcal{W} \otimes \mathcal{Z})$ be the state defined as

$$\rho = \frac{1}{m} \mathrm{Tr}_{\mathcal{Y}}\big(\mathrm{vec}(AV)\,\mathrm{vec}(AV)^*\big), \tag{8.329}$$

and let $\omega \in \mathrm{D}(\mathcal{Z})$ denote the completely mixed state on $\mathcal{Z}$. There exists a channel $\Xi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ such that

$$\begin{aligned} \mathrm{F}\left(\frac{1}{m}\mathrm{vec}(\mathbb{1}_{\mathcal{Z}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{Z}})^*,\, \frac{1}{m}(\Xi\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})\big(\mathrm{vec}(V)\,\mathrm{vec}(V)^*\big)\right) \\ \geq \mathrm{F}\big(\rho, \mathrm{Tr}_{\mathcal{Z}}(\rho) \otimes \omega\big). \end{aligned} \tag{8.330}$$

Proof. Let $\mathcal{V}$ be a complex Euclidean space of sufficiently large dimension that the inequalities $\dim(\mathcal{V}) \geq \dim(\mathcal{W})$ and $\dim(\mathcal{V} \otimes \mathcal{Z}) \geq \dim(\mathcal{Y})$ hold, and let $B \in \mathrm{L}(\mathcal{W}, \mathcal{V})$ be an operator such that $\mathrm{Tr}_{\mathcal{V}}\big(\mathrm{vec}(B)\,\mathrm{vec}(B)^*\big) = \mathrm{Tr}_{\mathcal{Z}}(\rho)$. For the vector

$$u = \frac{1}{\sqrt{m}}\mathrm{vec}(B \otimes \mathbb{1}_{\mathcal{Z}}) \in (\mathcal{V} \otimes \mathcal{Z}) \otimes (\mathcal{W} \otimes \mathcal{Z}), \tag{8.331}$$

one has that $\mathrm{Tr}_{\mathcal{V}\otimes\mathcal{Z}}(uu^*) = \mathrm{Tr}_{\mathcal{Z}}(\rho) \otimes \omega$. It is evident that the vector

$$v = \frac{1}{\sqrt{m}}\mathrm{vec}(AV) \in \mathcal{Y} \otimes \mathcal{W} \otimes \mathcal{Z} \tag{8.332}$$

satisfies $\mathrm{Tr}_{\mathcal{Y}}(vv^*) = \rho$, so it follows by Uhlmann's theorem (Theorem 3.23) that there exists an isometry $W \in \mathrm{U}(\mathcal{Y}, \mathcal{V} \otimes \mathcal{Z})$ such that

$$\mathrm{F}\big(\rho, \mathrm{Tr}_{\mathcal{Z}}(\rho) \otimes \omega\big) = \big|\big\langle u, (W \otimes \mathbb{1}_{\mathcal{W}\otimes\mathcal{Z}})v\big\rangle\big|. \tag{8.333}$$

Now define a channel $\Xi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ as

$$\Xi(Y) = \mathrm{Tr}_{\mathcal{V}}(WYW^*) \tag{8.334}$$

for every $Y \in \mathrm{L}(\mathcal{Y})$. By the monotonicity of the fidelity under partial tracing (which is a special case of Theorem 3.30), one has

$$\begin{aligned}
&\mathrm{F}\big(\rho, \mathrm{Tr}_{\mathcal{Z}}(\rho) \otimes \omega\big) \\
&\quad = \mathrm{F}\big(uu^*, (W \otimes \mathbb{1}_{\mathcal{W}\otimes\mathcal{Z}})vv^*(W \otimes \mathbb{1}_{\mathcal{W}\otimes\mathcal{Z}})^*\big) \\
&\quad \leq \mathrm{F}\Bigg(\mathrm{Tr}_{\mathcal{V}}\big(\mathrm{Tr}_{\mathcal{W}}(uu^*)\big), \frac{1}{m}(\Xi\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})\big(\mathrm{vec}(V)\,\mathrm{vec}(V)^*\big)\Bigg) \\
&\quad = \mathrm{F}\Bigg(\frac{1}{m}\,\mathrm{vec}(\mathbb{1}_{\mathcal{Z}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{Z}})^*, \frac{1}{m}(\Xi\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(\mathrm{vec}(V)\,\mathrm{vec}(V)^*)\Bigg).
\end{aligned} \tag{8.335}$$

The channel Ξ therefore satisfies the requirement of the lemma. □

Two additional lemmas needed for the quantum capacity theorem

The two lemmas that follow represent technical facts that will be utilized in the proof of the quantum capacity theorem. The first lemma concerns the approximation of one isometry by another isometry that meets certain spectral requirements, and the second lemma is a general fact regarding Haar measure.

Lemma 8.55. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{W}$ be complex Euclidean spaces, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y}\otimes\mathcal{W})$ be an isometry, let $\Lambda \in \mathrm{Proj}(\mathcal{Y})$ and $\Pi \in \mathrm{Proj}(\mathcal{W})$ be projection operators, and let $\varepsilon \in (0, 1/4)$ be a positive real number. Also let $n = \dim(\mathcal{X})$, and assume that the constraints*

$$\big\langle \Lambda \otimes \Pi, AA^* \big\rangle \geq (1-\varepsilon)n \tag{8.336}$$

and

$$2\,\mathrm{rank}(\Pi) \leq \dim(\mathcal{W}) \tag{8.337}$$

are satisfied. There exists an isometry $B \in \mathrm{U}(\mathcal{X}, \mathcal{Y}\otimes\mathcal{W})$ such that

1. $\|A - B\|_2 < 3\varepsilon^{1/4}\sqrt{n}$,
2. $\mathrm{Tr}_{\mathcal{W}}(BB^*) \leq 4\Lambda\,\mathrm{Tr}_{\mathcal{W}}(AA^*)\Lambda$, *and*
3. $\mathrm{rank}(\mathrm{Tr}_{\mathcal{Y}}(BB^*)) \leq 2\,\mathrm{rank}(\Pi)$.

Proof. By means of the singular value theorem, one may write

$$(\Lambda \otimes \Pi)A = \sum_{k=1}^{n} s_k u_k x_k^* \tag{8.338}$$

for an orthonormal basis $\{x_1, \ldots, x_n\}$ of $\mathcal{X}$, an orthonormal set $\{u_1, \ldots, u_n\}$ of vectors in $\mathcal{Y} \otimes \mathcal{W}$, and a collection $\{s_1, \ldots, s_n\} \subset [0, 1]$ of nonnegative real numbers. It holds that

$$\sum_{k=1}^{n} s_k^2 = \left\langle \Lambda \otimes \Pi, AA^* \right\rangle \geq (1 - \varepsilon)n. \tag{8.339}$$

Define $\Gamma \subseteq \{1, \ldots, n\}$ as

$$\Gamma = \left\{ k \in \{1, \ldots, n\} \,:\, s_k^2 \geq 1 - \sqrt{\varepsilon} \right\}, \tag{8.340}$$

and observe the inequalities

$$\begin{aligned} (1 - \varepsilon)n &\leq \sum_{k=1}^{n} s_k^2 \\ &\leq |\Gamma| + (n - |\Gamma|)\left(1 - \sqrt{\varepsilon}\right) = \left(1 - \sqrt{\varepsilon}\right)n + \sqrt{\varepsilon}|\Gamma|, \end{aligned} \tag{8.341}$$

from which it follows that

$$|\Gamma| \geq \left(1 - \sqrt{\varepsilon}\right) n > \frac{n}{2}. \tag{8.342}$$

Let $f : \{1, \ldots, n\} \backslash \Gamma \to \Gamma$ be any one-to-one function, and let $W \in \mathrm{U}(\mathcal{W})$ be any unitary operator satisfying $\Pi W \Pi = 0$; the existence of such an operator W follows from the assumption $2 \operatorname{rank}(\Pi) \leq \dim(\mathcal{W})$. Finally, define an isometry $B \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ as follows:

$$B = \sum_{k \in \Gamma} u_k x_k^* + \sum_{k \in \{1, \ldots, n\} \backslash \Gamma} (\mathbb{1}_{\mathcal{Y}} \otimes W) u_{f(k)} x_k^*. \tag{8.343}$$

It remains to prove that B has the properties required by the statement of the lemma.

First, it will be verified that B is indeed an isometry. The set $\{u_k \,:\, k \in \Gamma\}$ is evidently orthonormal, as is the set

$$\{(\mathbb{1}_{\mathcal{Y}} \otimes W) u_{f(k)} \,:\, k \in \{1, \ldots, n\} \backslash \Gamma\}. \tag{8.344}$$

Moreover, for every $k \in \Gamma$, it must hold that $s_k > 0$, and therefore

$$u_k \in \operatorname{im}\big((\Lambda \otimes \Pi)A\big) \subseteq \operatorname{im}\big(\Lambda \otimes \Pi\big), \tag{8.345}$$

which implies that $u_k = (\mathbb{1}_{\mathcal{Y}} \otimes \Pi)u_k$. For an arbitrary choice of $j, k \in \Gamma$, one therefore has

$$\begin{aligned} \big\langle u_j, (\mathbb{1}_{\mathcal{Y}} \otimes W)u_k \big\rangle &= \big\langle (\mathbb{1}_{\mathcal{Y}} \otimes \Pi)u_j, (\mathbb{1}_{\mathcal{Y}} \otimes W\Pi)u_k \big\rangle \\ &= \big\langle u_j, (\mathbb{1}_{\mathcal{Y}} \otimes \Pi W \Pi)u_k \big\rangle = 0, \end{aligned} \tag{8.346}$$

which implies that the set

$$\{u_k \,:\, k \in \Gamma\} \cup \big\{(\mathbb{1}_{\mathcal{Y}} \otimes W)u_{f(k)} \,:\, k \in \{1, \ldots, n\} \backslash \Gamma\big\} \tag{8.347}$$

is an orthonormal set. This implies that B is an isometry.

Next, observe that

$$\|A - B\|_2 \leq \|A - (\Lambda \otimes \Pi)A\|_2 + \|(\Lambda \otimes \Pi)A - B\|_2. \tag{8.348}$$

The first term in this expression is bounded as

$$\|A - (\Lambda \otimes \Pi)A\|_2 = \sqrt{\big\langle \mathbb{1} - \Lambda \otimes \Pi, AA^* \big\rangle} \leq \sqrt{\varepsilon n}. \tag{8.349}$$

For the second term, it holds that

$$\begin{aligned} \big\|(\Lambda \otimes \Pi)A - B\big\|_2^2 &= \sum_{k \in \Gamma} (s_k - 1)^2 + \sum_{k \in \{1,\ldots,n\} \backslash \Gamma} \big(s_k^2 + 1\big) \\ &= n + \sum_{k=1}^{n} s_k^2 - 2 \sum_{k \in \Gamma} s_k. \end{aligned} \tag{8.350}$$

Consequently, one finds that

$$\big\|(\Lambda \otimes \Pi)A - B\big\|_2^2 \leq 2n - 2\big(1 - \sqrt{\varepsilon}\big)^{\frac{3}{2}} n < 4n\sqrt{\varepsilon}. \tag{8.351}$$

It follows that

$$\|A - B\|_2 < 3\varepsilon^{1/4}\sqrt{n}, \tag{8.352}$$

so that the first requirement on B is fulfilled.

The second requirement on B may be verified as follows:

$$\begin{aligned} \operatorname{Tr}_{\mathcal{W}}\big(BB^*\big) &\leq 2 \sum_{k \in \Gamma} \operatorname{Tr}_{\mathcal{W}}\big(u_k u_k^*\big) \\ &\leq \frac{2}{1 - \sqrt{\varepsilon}} \operatorname{Tr}_{\mathcal{W}}\big((\Lambda \otimes \Pi)AA^*(\Lambda \otimes \Pi)\big) \leq 4\Lambda \operatorname{Tr}_{\mathcal{W}}(AA^*)\Lambda. \end{aligned} \tag{8.353}$$

Finally, to verify that the third requirement on B is satisfied, one may again use the observation that $(\mathbb{1} \otimes \Pi)u_k = u_k$, which implies that

$$\operatorname{im}(\operatorname{Tr}_{\mathcal{Y}}(u_k u_k^*)) \subseteq \operatorname{im}(\Pi), \tag{8.354}$$

for each $k \in \Gamma$. As

$$\operatorname{Tr}_{\mathcal{Y}}(BB^*) = \sum_{k\in\Gamma} \operatorname{Tr}_{\mathcal{Y}}(u_k u_k^*) + \sum_{k\in\{1,\ldots,n\}\setminus\Gamma} W(\operatorname{Tr}_{\mathcal{Y}}(u_{f(k)} u_{f(k)}^*))W^*, \tag{8.355}$$

it follows that

$$\operatorname{im}\big(\operatorname{Tr}_{\mathcal{Y}}(BB^*)\big) \subseteq \operatorname{im}(\Pi) + \operatorname{im}(W\Pi) \tag{8.356}$$

and therefore

$$\operatorname{rank}\big(\operatorname{Tr}_{\mathcal{Y}}(BB^*)\big) \leq 2\operatorname{rank}(\Pi), \tag{8.357}$$

as required. □

Lemma 8.56. *Let $\mathcal{X}$, $\mathcal{W}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $m = \dim(\mathcal{Z})$ and $n = \dim(\mathcal{X})$, and assume $m \leq n$. For every choice of an isometry $V \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$ and every operator $Z \in \mathrm{L}(\mathcal{W} \otimes \mathcal{X})$, it holds that*

$$\int \big\| \big(\mathbb{1}_{\mathcal{W}} \otimes V^* U^*\big) Z \big(\mathbb{1}_{\mathcal{W}} \otimes UV\big) \big\|_1 \, \mathrm{d}\eta(U) \leq \frac{m}{n} \|Z\|_1 \tag{8.358}$$

for η denoting the Haar measure on $\mathrm{U}(\mathcal{X})$.

Proof. Let

$$\big\{W_1, \ldots, W_{n^2}\big\} \subset \mathrm{U}(\mathcal{X}) \tag{8.359}$$

be any collection of unitary operators for which it holds that the completely depolarizing channel $\Omega \in \mathrm{C}(\mathcal{X})$ is given by

$$\Omega(X) = \frac{1}{n^2} \sum_{k=1}^{n^2} W_k X W_k^* \tag{8.360}$$

for all $X \in \mathrm{L}(\mathcal{X})$. (Such a collection may, for instance, be derived from the discrete Weyl operators defined in Section 4.1.2.) Define $\mathcal{Y} = \mathbb{C}^{n^2}$, and define a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Z} \otimes \mathcal{Y})$ as

$$\Phi(X) = \frac{1}{nm} \sum_{k=1}^{n^2} V^* W_k^* X W_k V \otimes E_{k,k} \tag{8.361}$$

for every $X \in \mathrm{L}(\mathcal{X})$. The fact that Φ is a channel follows from Corollary 2.27 together with the calculation

$$\frac{1}{nm} \sum_{k=1}^{n^2} W_k V V^* W_k^* = \frac{n}{m} \Omega(VV^*) = \mathbb{1}_{\mathcal{X}}. \tag{8.362}$$

Next, by the right unitary invariance of the Haar measure, it holds that

$$\begin{aligned} &\int \| \, (\mathbb{1}_{\mathcal{W}} \otimes V^* U^*) Z (\mathbb{1}_{\mathcal{W}} \otimes UV) \, \|_1 \, \mathrm{d}\eta(U) \\ &\quad = \int \| \, (\mathbb{1}_{\mathcal{W}} \otimes V^* W_k^* U^*) Z (\mathbb{1}_{\mathcal{W}} \otimes U W_k V) \, \|_1 \, \mathrm{d}\eta(U) \end{aligned} \tag{8.363}$$

for every choice of $k \in \{1, \ldots, n^2\}$, and therefore

$$\begin{aligned} &\int \| \, (\mathbb{1}_{\mathcal{W}} \otimes UV)^* Z (\mathbb{1}_{\mathcal{W}} \otimes UV) \, \|_1 \, \mathrm{d}\eta(U) \\ &= \frac{1}{n^2} \sum_{k=1}^{n^2} \int \Big\| \, (\mathbb{1}_{\mathcal{W}} \otimes U W_k V)^* Z (\mathbb{1}_{\mathcal{W}} \otimes U W_k V) \, \Big\|_1 \, \mathrm{d}\eta(U) \\ &= \frac{1}{n^2} \int \left\| \sum_{k=1}^{n^2} (\mathbb{1}_{\mathcal{W}} \otimes U W_k V)^* Z (\mathbb{1}_{\mathcal{W}} \otimes U W_k V) \otimes E_{k,k} \right\|_1 \mathrm{d}\eta(U) \\ &= \frac{m}{n} \int \| \, (\mathbb{1}_{\mathrm{L}(\mathcal{W})} \otimes \Phi) \big((\mathbb{1}_{\mathcal{W}} \otimes U) Z (\mathbb{1}_{\mathcal{W}} \otimes U^*) \big) \, \|_1 \, \mathrm{d}\eta(U). \end{aligned} \tag{8.364}$$

As the trace norm is non-increasing under the action of channels, as well as unitarily invariant, it follows that

$$\begin{aligned} &\| \, (\mathbb{1}_{\mathcal{W}} \otimes UV)^* Z (\mathbb{1}_{\mathcal{W}} \otimes UV) \, \|_1 \, \mathrm{d}\eta(U) \\ &\quad \leq \frac{m}{n} \int \| \, (\mathbb{1}_{\mathcal{W}} \otimes U) Z (\mathbb{1}_{\mathcal{W}} \otimes U^*) \, \|_1 \, \mathrm{d}\eta(U) = \frac{m}{n} \| Z \|_1, \end{aligned} \tag{8.365}$$

which completes the proof. □

The quantum capacity theorem

As the following theorem establishes, the entanglement-generation capacity of a given channel is always at least as large as the coherent information of the completely mixed state through that channel. This fact, which will be generalized to arbitrary states in place of the completely mixed state in a corollary to the theorem, lies at the heart of the proof of the quantum capacity theorem.

Theorem 8.57. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. The entanglement generation capacity of Φ is lower-bounded by the coherent information of the completely mixed state $\omega \in \mathrm{D}(\mathcal{X})$ through Φ:*

$$\mathrm{I}_{\mathrm{C}}(\omega\,;\,\Phi) \leq \mathrm{Q}_{\mathrm{EG}}(\Phi). \tag{8.366}$$

Proof. Let $\mathcal{W}$ be a complex Euclidean space such that

$$\dim(\mathcal{W}) = 2\dim(\mathcal{X} \otimes \mathcal{Y}), \tag{8.367}$$

and let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ be an isometry for which

$$\Phi(X) = \mathrm{Tr}_{\mathcal{W}}\big(AXA^*\big) \tag{8.368}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Define a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{W})$ as

$$\Psi(X) = \mathrm{Tr}_{\mathcal{Y}}\big(AXA^*\big) \tag{8.369}$$

for all $X \in \mathrm{L}(\mathcal{X})$, so that Ψ is complementary to Φ. It therefore holds that

$$\mathrm{I}_{\mathrm{C}}(\omega\,;\,\Phi) = \mathrm{H}(\Phi(\omega)) - \mathrm{H}(\Psi(\omega)). \tag{8.370}$$

The theorem is vacuous in the case that $\mathrm{I}_{\mathrm{C}}(\omega\,;\,\Phi) \leq 0$, so hereafter it will be assumed that $\mathrm{I}_{\mathrm{C}}(\omega\,;\,\Phi)$ is positive. To prove the theorem, it suffices to demonstrate that every positive real number smaller than $\mathrm{I}_{\mathrm{C}}(\omega\,;\,\Phi)$ is an achievable rate for entanglement generation through Φ. Toward this goal, assume that an arbitrary positive real number α satisfying $\alpha < \mathrm{I}_{\mathrm{C}}(\omega\,;\,\Phi)$ has been fixed, and that $\varepsilon > 0$ is a positive real number chosen to be sufficiently small so that the inequality

$$\alpha < \mathrm{I}_{\mathrm{C}}(\omega\,;\,\Phi) - 2\varepsilon(\mathrm{H}(\Phi(\omega)) + \mathrm{H}(\Psi(\omega))) \tag{8.371}$$

is satisfied. The remainder of the proof is devoted to proving that α is an achievable rate for entanglement generation through Φ.

Consider an arbitrary positive integer $n \geq 1/\alpha$, and let m be any positive integer such that $m \leq \alpha n$. Also let $\Gamma = \{0, 1\}$ denote the binary alphabet, and let $\mathcal{Z} = \mathbb{C}^{\Gamma}$. The task in which a state having high fidelity with the maximally entangled state

$$2^{-m}\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)^* \tag{8.372}$$

is established between a sender and receiver through the channel $\Phi^{\otimes n}$ is to be considered. For any choice of an isometry $W \in \mathrm{U}(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n})$ and a channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$, the state

$$2^{-m} \big(\Xi \Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big) \big(\mathrm{vec}(W)\, \mathrm{vec}(W)^*\big) \tag{8.373}$$

may be established through the channel $\Phi^{\otimes n}$, so one may aim to prove that there exists a choice of Ξ and W for which the fidelity between the states (8.372) and (8.373) is high.

It is helpful at this point to let $A_n \in \mathrm{U}(\mathcal{X}^{\otimes n}, \mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n})$ be the isometry defined by the equation

$$\begin{aligned} &\big(y_1 \otimes \cdots \otimes y_n \otimes w_1 \otimes \cdots \otimes w_n\big)^* A_n (x_1 \otimes \cdots \otimes x_n) \\ &\qquad = \prod_{k=1}^{n} (y_k \otimes w_k)^* A x_k \end{aligned} \tag{8.374}$$

holding for every choice of vectors $x_1, \ldots, x_n \in \mathcal{X}$, $y_1, \ldots, y_n \in \mathcal{Y}$, and $w_1, \ldots, w_n \in \mathcal{W}$. In effect, A_n is equivalent to $A^{\otimes n}$, except that the tensor factors in its output space have been permuted, so that the output space becomes $\mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n}$ rather than $(\mathcal{Y} \otimes \mathcal{W})^{\otimes n}$. It may be noted that

$$\Phi^{\otimes n}(X) = \mathrm{Tr}_{\mathcal{W}^{\otimes n}} \big(A_n X A_n^*\big) \quad \text{and} \quad \Psi^{\otimes n}(X) = \mathrm{Tr}_{\mathcal{Y}^{\otimes n}} \big(A_n X A_n^*\big) \tag{8.375}$$

for every $X \in \mathrm{L}(\mathcal{X}^{\otimes n})$. Under the assumption that the decoding channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ has been selected optimally, Lemma 8.54 implies that the fidelity between the states (8.372) and (8.373) is lower-bounded by

$$\mathrm{F}\big(\rho\,,\, \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\rho) \otimes \omega_{\mathcal{Z}}^{\otimes m}\big) \tag{8.376}$$

for $\rho \in \mathrm{D}(\mathcal{W}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ being defined as

$$\rho = 2^{-m}\, \mathrm{Tr}_{\mathcal{Y}^{\otimes n}} \big(\mathrm{vec}(A_n W)\, \mathrm{vec}(A_n W)^*\big) \tag{8.377}$$

and for $\omega_{\mathcal{Z}} \in \mathrm{D}(\mathcal{Z})$ denoting the completely mixed state on $\mathcal{Z}$.

The probabilistic method will be employed to prove the existence of an isometry W for which the expression (8.376) is close to 1, provided that n is sufficiently large. In particular, one may fix $V \in \mathrm{U}(\mathcal{Z}^{\otimes m}, \mathcal{X}^{\otimes n})$ to be an arbitrary isometry, and let $W = UV$ for U chosen at random with respect to the Haar measure on $\mathrm{U}(\mathcal{X}^{\otimes n})$. The analysis that follows demonstrates that, for an operator W chosen in this way, one expects the quantity (8.376) to be

close to 1, for sufficiently large n, which proves the existence of a choice of W for which this is true.

Let $k = \dim(\mathcal{X})$ and define $\xi \in \mathrm{D}(\mathcal{W}^{\otimes n} \otimes \mathcal{X}^{\otimes m})$ as

$$\xi = \frac{1}{k^n} \operatorname{Tr}_{\mathcal{Y}^{\otimes n}} \bigl(\operatorname{vec}(A_n) \operatorname{vec}(A_n)^*\bigr). \tag{8.378}$$

Also define $\rho_U \in \mathrm{D}(\mathcal{W}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$ as

$$\rho_U = \frac{1}{2^m} \operatorname{Tr}_{\mathcal{Y}^{\otimes n}} \bigl(\operatorname{vec}(A_n U V) \operatorname{vec}(A_n U V)^*\bigr), \tag{8.379}$$

for each unitary operator $U \in \mathrm{U}(\mathcal{X}^{\otimes n})$, and observe that

$$\rho_U = \frac{k^n}{2^m} \bigl(\mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes V^{\mathsf{T}} U^{\mathsf{T}}\bigr) \xi \bigl(\mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes V^{\mathsf{T}} U^{\mathsf{T}}\bigr)^*. \tag{8.380}$$

For the isometry $W = UV$, the fidelity between the states (8.372) and (8.373) is lower-bounded by

$$\mathrm{F}\bigl(\rho_U \,,\, \operatorname{Tr}_{\mathcal{Z}^{\otimes m}}(\rho_U) \otimes \omega_{\mathcal{Z}}^{\otimes m}\bigr), \tag{8.381}$$

for a suitable choice of the decoding channel Ξ.

Let $\Lambda_{n,\varepsilon} \in \mathrm{Proj}(\mathcal{Y}^{\otimes n})$ and $\Pi_{n,\varepsilon} \in \mathrm{Proj}(\mathcal{W}^{\otimes n})$ be the projection operators onto the ε-strongly typical subspaces of $\mathcal{Y}^{\otimes n}$ and $\mathcal{W}^{\otimes n}$, with respect to any fixed choice of spectral decompositions of the operators $\Phi(\omega)$ and $\Psi(\omega)$, respectively. By Lemma 8.35, one has the inequalities

$$\begin{aligned} &\frac{1}{k^n} \bigl\langle \Lambda_{n,\varepsilon} \otimes \mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes \mathbb{1}_{\mathcal{X}}^{\otimes n}, \operatorname{vec}(A_n) \operatorname{vec}(A_n)^* \bigr\rangle \\ &\qquad = \bigl\langle \Lambda_{n,\varepsilon}, \bigl(\Phi(\omega)\bigr)^{\otimes n} \bigr\rangle > 1 - \zeta_{n,\varepsilon}, \\ &\frac{1}{k^n} \bigl\langle \mathbb{1}_{\mathcal{Y}}^{\otimes n} \otimes \Pi_{n,\varepsilon} \otimes \mathbb{1}_{\mathcal{X}}^{\otimes n}, \operatorname{vec}(A_n) \operatorname{vec}(A_n)^* \bigr\rangle \\ &\qquad = \bigl\langle \Pi_{n,\varepsilon}, \bigl(\Psi(\omega)\bigr)^{\otimes n} \bigr\rangle > 1 - \zeta_{n,\varepsilon} \end{aligned} \tag{8.382}$$

for

$$\zeta_{n,\varepsilon} = K \exp(-\delta n \varepsilon^2), \tag{8.383}$$

where $K \geq 1$ and $\delta > 0$ are positive real numbers that are independent of n and ε. It follows that

$$\frac{1}{k^n} \bigl\langle \Lambda_{n,\varepsilon} \otimes \Pi_{n,\varepsilon} \otimes \mathbb{1}_{\mathcal{X}}^{\otimes n}, \operatorname{vec}(A_n) \operatorname{vec}(A_n)^* \bigr\rangle > 1 - 2\zeta_{n,\varepsilon}, \tag{8.384}$$

which is equivalent to

$$\left\langle \Lambda_{n,\varepsilon} \otimes \Pi_{n,\varepsilon}, A_n A_n^* \right\rangle \geq (1 - 2\zeta_{n,\varepsilon})\, k^n. \tag{8.385}$$

If n is sufficiently large so that $\zeta_{n,\varepsilon} < 1/4$, it follows by Lemma 8.55 that there exists an isometry $B_n \in \mathrm{U}(\mathcal{X}^{\otimes n}, \mathcal{Y}^{\otimes n} \otimes \mathcal{W}^{\otimes n})$ satisfying the conditions

$$\begin{aligned} \left\| A_n - B_n \right\|_2 &\leq 3\, \zeta_{n,\varepsilon}^{1/4}\, k^{n/2}, \\ \mathrm{Tr}_{\mathcal{W}^{\otimes n}} \left(B_n B_n^* \right) &\leq 4 \Lambda_{n,\varepsilon}\, \mathrm{Tr}_{\mathcal{W}^{\otimes n}} \left(A_n A_n^* \right) \Lambda_{n,\varepsilon}, \end{aligned} \tag{8.386}$$

and

$$\mathrm{rank}\left(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}} \left(B_n B_n^* \right) \right) \leq 2\, \mathrm{rank}(\Pi_{n,\varepsilon}). \tag{8.387}$$

By Proposition 8.36, the third condition implies that

$$\mathrm{rank}\left(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}} \left(B_n B_n^* \right) \right) \leq 2^{n(1+\varepsilon)\, \mathrm{H}(\Psi(\omega))+1}. \tag{8.388}$$

Using the second condition, together with Corollary 8.34, one obtains

$$\begin{aligned} &\mathrm{Tr}\left(\left(\frac{1}{k^n} \mathrm{Tr}_{\mathcal{W}^{\otimes n}} \left(B_n B_n^* \right) \right)^2 \right) \\ &\qquad \leq \mathrm{Tr}\left(\left(\frac{4}{k^n} \Lambda_{n,\varepsilon}\, \mathrm{Tr}_{\mathcal{W}^{\otimes n}} \left(A_n A_n^* \right) \Lambda_{n,\varepsilon} \right)^2 \right) \\ &\qquad = 16\, \mathrm{Tr}\Big(\left(\Lambda_{n,\varepsilon} \Phi(\omega)^{\otimes n} \Lambda_{n,\varepsilon} \right)^2 \Big) \\ &\qquad \leq 2^{-n(1-\varepsilon)\, \mathrm{H}(\Phi(\omega))+4}. \end{aligned} \tag{8.389}$$

Finally, define

$$\sigma = \frac{1}{k^n} \mathrm{Tr}_{\mathcal{Y}^{\otimes n}} \left(\mathrm{vec}(B_n)\, \mathrm{vec}(B_n)^* \right), \tag{8.390}$$

and also define

$$\begin{aligned} \tau_U &= \frac{1}{2^m} \mathrm{Tr}_{\mathcal{Y}^{\otimes n}} \left(\mathrm{vec}(B_n U V)\, \mathrm{vec}(B_n U V)^* \right) \\ &= \frac{k^n}{2^m} \left(\mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes V^{\mathsf{T}} U^{\mathsf{T}} \right) \sigma \left(\mathbb{1}_{\mathcal{W}}^{\otimes n} \otimes V^{\mathsf{T}} U^{\mathsf{T}} \right)^* \end{aligned} \tag{8.391}$$

for each $U \in \mathrm{U}(\mathcal{X}^{\otimes n})$. It holds that

$$\begin{aligned}
&\Big\| \rho_U - \mathrm{Tr}_{\mathcal{Z}}(\rho_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big\|_1 \\
&\qquad \leq \big\| \rho_U - \tau_U \big\|_1 + \Big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big\|_1 \\
&\qquad\qquad + \Big\| \big(\mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\rho_U)\big) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big\|_1 \\
&\qquad \leq \Big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \Big\|_1 + 2 \big\| \rho_U - \tau_U \big\|_1,
\end{aligned} \tag{8.392}$$

and it remains to consider the average value of the two terms in the final expression of this inequality. When considering the average value of first term in the final expression of (8.392), it may be noted that

$$\begin{aligned}
\mathrm{rank}\Big(\tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m}\Big) &\leq \mathrm{rank}\Big(\mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m}\Big) \\
&\leq 2^m \ \mathrm{rank}\big(\mathrm{Tr}_{\mathcal{Y}^{\otimes n}}\big(B_n B_n^*\big)\big) \leq 2^{n(1+\varepsilon)\,\mathrm{H}(\Psi(\omega))+m+1}
\end{aligned} \tag{8.393}$$

and

$$\mathrm{Tr}(\sigma^2) = \mathrm{Tr}\left(\left(\frac{1}{k^n}\,\mathrm{Tr}_{\mathcal{W}^{\otimes n}}\big(B_n B_n^*\big)\right)^2\right) \leq 2^{-n(1-\varepsilon)\,\mathrm{H}(\Phi(\omega))+4}. \tag{8.394}$$

Making use of Lemma 8.53, it therefore follows that

$$\begin{aligned}
&\int \big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \big\|_1^2 \,\mathrm{d}\eta(U) \\
&\qquad \leq 2^{n(1+\varepsilon)\,\mathrm{H}(\Psi(\omega))+m+1} \int \big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \big\|_2^2 \,\mathrm{d}\eta(U) \\
&\qquad \leq 2^{n((1+\varepsilon)\,\mathrm{H}(\Psi(\omega))-(1-\varepsilon)\,\mathrm{H}(\Phi(\omega)))+m+5} \\
&\qquad = 2^{-n(\mathrm{I_c}(\omega;\Phi)-2\varepsilon(\mathrm{H}(\Phi(\omega))+\mathrm{H}(\Psi(\omega))))+m+5}.
\end{aligned} \tag{8.395}$$

By the assumption (8.371), along with the assumption that $m \leq \alpha n$, one has that this quantity approaches 0 in the limit as n approaches infinity. It therefore holds (by Jensen's inequality) that the quantity

$$\int \big\| \tau_U - \mathrm{Tr}_{\mathcal{Z}^{\otimes m}}(\tau_U) \otimes \omega_{\mathcal{Z}}^{\otimes m} \big\|_1 \,\mathrm{d}\eta(U) \tag{8.396}$$

also approaches 0 in the limit as n approaches infinity. The average value of the second term in the final expression of (8.392) may be upper-bounded as

$$\begin{aligned}
&\int \|\rho_U - \tau_U\|_1 \, \mathrm{d}\eta(U) \\
&= \frac{k^n}{2^m} \int \left\| \left(\mathbb{1}_{\mathcal{Y}^{\otimes n}} \otimes V^{\mathsf{T}} U^{\mathsf{T}}\right)(\xi - \sigma)\left(\mathbb{1}_{\mathcal{Y}^{\otimes n}} \otimes V^{\mathsf{T}} U^{\mathsf{T}}\right)^* \right\|_1 \mathrm{d}\eta(U) \\
&\leq \|\xi - \sigma\|_1 \leq \frac{1}{k^n} \left\| \operatorname{vec}(A_n) \operatorname{vec}(A_n)^* - \operatorname{vec}(B_n) \operatorname{vec}(B_n)^* \right\|_1 \\
&\leq \frac{2}{k^{n/2}} \left\| A_n - B_n \right\|_2 \leq 6\, \zeta_{n,\varepsilon}^{1/4}
\end{aligned} \tag{8.397}$$

by Lemma 8.56. Once again, this quantity approaches 0 in the limit as n approaches infinity. It follows that the entanglement generation capacity of Φ is at least α, which completes the proof. □

Corollary 8.58. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel, and let $\sigma \in \mathrm{D}(\mathcal{X})$ be a density operator. The quantum capacity of Φ is lower-bounded by the coherent information of σ through Φ:*

$$\mathrm{I}_{\mathrm{C}}(\sigma; \Phi) \leq \mathrm{Q}(\Phi). \tag{8.398}$$

Proof. Observe first that it is a consequence of Theorem 8.57 that

$$\mathrm{I}_{\mathrm{C}}(\omega_{\mathcal{V}}; \Phi) \leq \mathrm{Q}(\Phi) \tag{8.399}$$

for every nontrivial subspace $\mathcal{V} \subseteq \mathcal{X}$, where

$$\omega_{\mathcal{V}} = \frac{\Pi_{\mathcal{V}}}{\dim(\mathcal{V})} \tag{8.400}$$

is the state that is maximally mixed over the subspace $\mathcal{V}$. To verify that this is so, let $\mathcal{Z}$ be any complex Euclidean space with $\dim(\mathcal{Z}) = \dim(\mathcal{V})$, let $V \in \mathrm{U}(\mathcal{Z}, \mathcal{X})$ be an isometry such that $VV^* = \Pi_{\mathcal{V}}$, and define a channel $\Xi \in \mathrm{C}(\mathcal{Z}, \mathcal{Y})$ as

$$\Xi(Z) = \Phi(VZV^*) \tag{8.401}$$

for all $Z \in \mathrm{L}(\mathcal{Z})$. It is evident that $\mathrm{Q}(\Xi) \leq \mathrm{Q}(\Phi)$; the channel Φ emulates Ξ, so for every positive integer n it holds that $\Phi^{\otimes n}$ emulates every channel that can be emulated by $\Xi^{\otimes n}$. It follows that

$$\begin{aligned}
\mathrm{Q}(\Phi) \geq \mathrm{Q}(\Xi) &= \mathrm{Q}_{\mathrm{EG}}(\Xi) \geq \mathrm{I}_{\mathrm{C}}(\omega_{\mathcal{Z}}; \Xi) \\
&= \mathrm{I}_{\mathrm{C}}(V \omega_{\mathcal{Z}} V^*; \Phi) = \mathrm{I}_{\mathrm{C}}(\omega_{\mathcal{V}}; \Phi),
\end{aligned} \tag{8.402}$$

as claimed.

Now, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{W})$ be an isometry such that

$$\Phi(X) = \mathrm{Tr}_{\mathcal{W}}(AXA^*) \tag{8.403}$$

for all $X \in \mathrm{L}(\mathcal{X})$, for a suitable choice of a complex Euclidean space $\mathcal{W}$, and define a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{W})$ as

$$\Psi(X) = \mathrm{Tr}_{\mathcal{Y}}(AXA^*) \tag{8.404}$$

for all $X \in \mathrm{L}(\mathcal{X})$. It therefore holds that Ψ is complementary to Φ, so that

$$\mathrm{I}_{\mathrm{c}}(\sigma; \Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Psi(\sigma)). \tag{8.405}$$

Let

$$\sigma = \sum_{a \in \Sigma} p(a) x_a x_a^* \tag{8.406}$$

be a spectral decomposition of σ, and let

$$\omega_{n,\varepsilon} = \frac{\Lambda_{n,\varepsilon}}{\mathrm{Tr}(\Lambda_{n,\varepsilon})} \in \mathrm{D}(\mathcal{X}^{\otimes n}) \tag{8.407}$$

for each positive integer n and each positive real number $\varepsilon > 0$, for $\Lambda_{n,\varepsilon}$ denoting the projection onto the ε-strongly typical subspace of $\mathcal{X}^{\otimes n}$, with respect to the spectral decomposition (8.406).

Next, let $\varepsilon > 0$ be a positive real number, to be chosen arbitrarily. By Lemma 8.39, it holds that

$$\left| \lim_{n \to \infty} \frac{\mathrm{H}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon})\big)}{n} - \mathrm{H}(\Phi(\sigma)) \right| \leq (2\,\mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)))\varepsilon, \tag{8.408}$$

and therefore there must exist a positive integer n_0 such that, for all $n \geq n_0$, one has

$$\left| \frac{1}{n} \mathrm{H}\big(\Phi^{\otimes n}(\omega_{n,\varepsilon})\big) - \mathrm{H}(\Phi(\sigma)) \right| \leq (2\,\mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) + 1)\varepsilon. \tag{8.409}$$

Along similar lines, there must exist a positive integer n_1 such that, for all $n \geq n_1$, one has

$$\left| \frac{1}{n} \mathrm{H}\big(\Psi^{\otimes n}(\omega_{n,\varepsilon})\big) - \mathrm{H}(\Psi(\sigma)) \right| \leq (2\,\mathrm{H}(\sigma) + \mathrm{H}(\Psi(\sigma)) + 1)\varepsilon. \tag{8.410}$$

It follows that there must exist a positive integer n such that

$$\begin{aligned}\left|\frac{1}{n}\mathrm{I}_{\mathrm{C}}\bigl(\omega_{n,\varepsilon};\Phi^{\otimes n}\bigr) - \mathrm{I}_{\mathrm{C}}(\sigma;\Phi)\right| \\ \leq (4\,\mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) + \mathrm{H}(\Psi(\sigma)) + 2)\varepsilon.\end{aligned} \tag{8.411}$$

By the argument presented at the beginning of the proof, it holds that

$$\frac{\mathrm{I}_{\mathrm{C}}\bigl(\omega_{n,\varepsilon};\Phi^{\otimes n}\bigr)}{n} \leq \frac{\mathrm{Q}(\Phi^{\otimes n})}{n} = \mathrm{Q}(\Phi), \tag{8.412}$$

and therefore

$$\mathrm{Q}(\Phi) \geq \mathrm{I}_{\mathrm{C}}\bigl(\sigma;\Phi\bigr) - (4\,\mathrm{H}(\sigma) + \mathrm{H}(\Phi(\sigma)) + \mathrm{H}(\Psi(\sigma)) + 2)\varepsilon. \tag{8.413}$$

As ε has been chosen to be an arbitrary positive real number, it follows that

$$\mathrm{Q}(\Phi) \geq \mathrm{I}_{\mathrm{C}}\bigl(\sigma;\Phi\bigr), \tag{8.414}$$

which completes the proof. □

Finally, the quantum capacity theorem may be stated and proved.

Theorem 8.59 (Quantum capacity theorem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X},\mathcal{Y})$ be a channel. It holds that*

$$\mathrm{Q}(\Phi) = \lim_{n\to\infty} \frac{\mathrm{I}_{\mathrm{C}}(\Phi^{\otimes n})}{n} \tag{8.415}$$

Proof. For every positive integer n and every density operator $\sigma \in \mathrm{D}(\mathcal{X}^{\otimes n})$, one has

$$\mathrm{I}_{\mathrm{C}}(\sigma\,;\,\Phi^{\otimes n}) \leq \mathrm{Q}(\Phi^{\otimes n}) = n\,\mathrm{Q}(\Phi) \tag{8.416}$$

by Corollary 8.58. It therefore holds that

$$\mathrm{Q}(\Phi) \geq \lim_{n\to\infty} \frac{\mathrm{I}_{\mathrm{C}}(\Phi^{\otimes n})}{n}. \tag{8.417}$$

Now suppose that α is an achievable rate for entanglement generation through Φ. It will be proved that

$$\lim_{n\to\infty} \frac{\mathrm{I}_{\mathrm{C}}(\Phi^{\otimes n})}{n} \geq \alpha. \tag{8.418}$$

If $\alpha = 0$, then this inequality holds trivially, so it will be assumed hereafter that $\alpha > 0$.

Let $\varepsilon \in (0, 1/2)$ be chosen arbitrarily, and let $\Gamma = \{0,1\}$ and $\mathcal{Z} = \mathbb{C}^{\Gamma}$. As α is an achievable rate for entanglement generation through Φ, it holds that for all but finitely many positive integers n and all positive integers $m \leq \alpha n$ that there must exist a unit vector $u \in \mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m}$ and a channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$ such that

$$\mathrm{F}\Big(2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^*, \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(uu^*)\Big) > 1 - \varepsilon, \tag{8.419}$$

and therefore

$$\Big\| 2^{-m} \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m}) \operatorname{vec}(\mathbb{1}_{\mathcal{Z}}^{\otimes m})^* - \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(uu^*) \Big\|_1 < 2\sqrt{2\varepsilon} \tag{8.420}$$

by one of the Fuchs–van de Graaf inequalities (Theorem 3.36). For any unit vector $u \in \mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m}$ for which the inequality (8.420) holds, one concludes from the Fannes–Audenaert inequality (Theorem 5.28) that

$$\mathrm{H}\big(\big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(uu^*)\big) \leq 2\delta m + \mathrm{H}(\delta, 1-\delta) \tag{8.421}$$

and

$$m - \mathrm{H}\big(\Xi\Phi^{\otimes n}(\rho)\big) \leq \delta m + \mathrm{H}(\delta, 1-\delta), \tag{8.422}$$

for

$$\rho = \operatorname{Tr}_{\mathcal{Z}^{\otimes m}}(uu^*) \quad \text{and} \quad \delta = \sqrt{2\varepsilon}. \tag{8.423}$$

Together, these inequalities imply that

$$\mathrm{I}_{\mathrm{C}}\big(\rho; \Xi\Phi^{\otimes n}\big) \geq (1 - 3\delta)m - 2\,\mathrm{H}(\delta, 1-\delta), \tag{8.424}$$

and therefore

$$\mathrm{I}_{\mathrm{C}}\big(\rho; \Phi^{\otimes n}\big) \geq (1 - 3\delta)m - 2\,\mathrm{H}(\delta, 1-\delta) \geq (1-3\delta)m - 2 \tag{8.425}$$

by Proposition 8.15. Choosing $m = \lfloor \alpha n \rfloor \geq \alpha n - 1$, it follows that

$$\frac{\mathrm{I}_{\mathrm{C}}\big(\rho; \Phi^{\otimes n}\big)}{n} \geq (1 - 3\delta)\alpha - \frac{3}{n} \tag{8.426}$$

so that

$$\lim_{n\to\infty} \frac{\mathrm{I}_{\mathrm{C}}(\Phi^{\otimes n})}{n} \geq (1 - 3\delta)\alpha. \tag{8.427}$$

Because δ may be taken to be arbitrarily small by a suitable choice of ε, the inequality (8.418) follows, which completes the proof. □

8.3 Non-additivity and super-activation

Expressions for the classical and quantum capacities of a quantum channel are given by regularizations of the Holevo capacity and maximum coherent information,

$$\mathrm{C}(\Phi) = \lim_{n\to\infty} \frac{\chi(\Phi^{\otimes n})}{n} \quad \text{and} \quad \mathrm{Q}(\Psi) = \lim_{n\to\infty} \frac{\mathrm{I}_{\mathrm{c}}(\Psi^{\otimes n})}{n}, \tag{8.428}$$

as has been established by the Holevo–Schumacher–Westmoreland theorem and quantum capacity theorem (Theorems 8.30 and 8.59). Non-regularized analogues of these formulas do not hold, in general. In particular, the strict inequalities

$$\chi(\Phi \otimes \Phi) > 2\chi(\Phi) \quad \text{and} \quad \mathrm{I}_{\mathrm{c}}(\Psi \otimes \Psi) > 2\,\mathrm{I}_{\mathrm{c}}(\Psi) \tag{8.429}$$

hold for a suitable choice of channels Φ and Ψ, as is demonstrated in the subsections that follow. These examples reveal that the Holevo capacity does not coincide directly with the classical capacity, and likewise for the maximum coherent information and quantum capacity.

With respect to the Holevo capacity, the fact that a strict inequality may hold for some channels Φ in (8.429) will be demonstrated in Section 8.3.1, through the use of Theorem 7.53 from the previous chapter. The existence of such channels is far from obvious, and no explicit examples are known at the time of this book's writing—it is only the existence of such channels that is known. The now falsified conjecture that the equality

$$\chi(\Phi_0 \otimes \Phi_1) = \chi(\Phi_0) + \chi(\Phi_1) \tag{8.430}$$

should hold for all choices of channels Φ_0 and Φ_1 was known for some time as the *additivity conjecture*.

In contrast, it is not difficult to find an example of a channel Ψ for which a strict inequality in (8.429) holds. There are, in fact, very striking examples of channels that go beyond the demonstration of non-additivity of maximum coherent information. In particular, one may find channels Ψ_0 and Ψ_1 such that both Ψ_0 and Ψ_1 have zero quantum capacity, and therefore

$$\mathrm{I}_{\mathrm{c}}(\Psi_0) = \mathrm{I}_{\mathrm{c}}(\Psi_1) = 0, \tag{8.431}$$

but for which

$$\mathrm{I}_{\mathrm{c}}(\Psi_0 \otimes \Psi_1) > 0, \tag{8.432}$$

and therefore $\Psi_0 \otimes \Psi_1$ has nonzero quantum capacity. This phenomenon is known as *super-activation*, and is discussed in Section 8.3.2. From such a choice of channels Ψ_0 and Ψ_1, the construction of a channel Ψ for which the strict inequality (8.429) holds is possible.

8.3.1 Non-additivity of the Holevo capacity

The fact that there exists a channel Φ for which

$$\chi(\Phi \otimes \Phi) > 2\chi(\Phi) \tag{8.433}$$

is demonstrated below. The proof makes use of Theorem 7.53, together with two basic ideas: one concerns the *direct sum* of two channels, and the other is a construction that relates the minimum output entropy of a given channel to the Holevo capacity of a channel constructed from the one given.

Direct sums of channels and their minimum output entropy

The direct sum of two maps is defined as follows. (One may also consider direct sums of more than two maps, but it is sufficient for the needs of the present section to consider the case of just two maps.)

Definition 8.60. Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0 \in \mathrm{T}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{T}(\mathcal{X}_1, \mathcal{Y}_1)$ be maps. The direct sum of Φ_0 and Φ_1 is the map

$$\Phi_0 \oplus \Phi_1 \in \mathrm{T}(\mathcal{X}_0 \oplus \mathcal{X}_1, \mathcal{Y}_0 \oplus \mathcal{Y}_1) \tag{8.434}$$

defined as

$$(\Phi_0 \oplus \Phi_1) \begin{pmatrix} X_0 & \cdot \\ \cdot & X_1 \end{pmatrix} = \begin{pmatrix} \Phi_0(X_0) & 0 \\ 0 & \Phi_1(X_1) \end{pmatrix} \tag{8.435}$$

for every $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$. The dots in (8.435) indicate arbitrary operators in $\mathrm{L}(\mathcal{X}_1, \mathcal{X}_0)$ and $\mathrm{L}(\mathcal{X}_0, \mathcal{X}_1)$ that have no influence on the output of the map $\Phi_0 \oplus \Phi_1$.

The direct sum of two channels is also a channel, as is established by the following straightforward proposition.

Proposition 8.61. *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels. The direct sum of Φ_0 and Φ_1 is a channel: $\Phi_0 \oplus \Phi_1 \in \mathrm{C}(\mathcal{X}_0 \oplus \mathcal{X}_1, \mathcal{Y}_0 \oplus \mathcal{Y}_1)$.*

Proof. It is immediate from the definition of the direct sum of Φ_0 and Φ_1 that $\Phi_0 \oplus \Phi_1$ is trace-preserving, so it suffices to prove that $\Phi_0 \oplus \Phi_1$ is completely positive. As Φ_0 and Φ_1 are completely positive, Kraus representations

$$\Phi_0(X_0) = \sum_{a\in\Sigma} A_a X_0 A_a^* \quad \text{and} \quad \Phi_1(X_1) = \sum_{b\in\Gamma} B_b X_1 B_b^* \tag{8.436}$$

of these maps must exist. Through a direct computation, one may verify that

$$\begin{aligned}(\Phi_0 \oplus \Phi_1)(X) = &\sum_{a\in\Sigma} \begin{pmatrix} A_a & 0 \\ 0 & 0 \end{pmatrix} X \begin{pmatrix} A_a & 0 \\ 0 & 0 \end{pmatrix}^* \\ &+ \sum_{b\in\Gamma} \begin{pmatrix} 0 & 0 \\ 0 & B_b \end{pmatrix} X \begin{pmatrix} 0 & 0 \\ 0 & B_b \end{pmatrix}^*\end{aligned} \tag{8.437}$$

for all $X \in \mathrm{L}(\mathcal{X}_0 \oplus \mathcal{X}_1)$. It follows that $\Phi_0 \oplus \Phi_1$ is completely positive, as required. □

By Theorem 7.53, there exist channels Φ_0 and Φ_1 such that

$$\mathrm{H}_{\min}(\Phi_0 \otimes \Phi_1) < \mathrm{H}_{\min}(\Phi_0) + \mathrm{H}_{\min}(\Phi_1). \tag{8.438}$$

It is possible to obtain, from this fact, an example of a single channel Φ such that

$$\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi). \tag{8.439}$$

The following corollary (to Theorem 7.53) establishes that this is so.

Corollary 8.62. *There exists a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, for some choice of complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, such that*

$$\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi). \tag{8.440}$$

Proof. By Theorem 7.53, there exist complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$ and channels $\Phi_0, \Phi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ such that

$$\mathrm{H}_{\min}(\Phi_0 \otimes \Phi_1) < \mathrm{H}_{\min}(\Phi_0) + \mathrm{H}_{\min}(\Phi_1). \tag{8.441}$$

Assume that such a choice of channels has been fixed for the remainder of the proof. Let $\sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X})$ be density operators satisfying

$$\mathrm{H}(\Phi_0(\sigma_0)) = \mathrm{H}_{\min}(\Phi_0) \quad \text{and} \quad \mathrm{H}(\Phi_1(\sigma_1)) = \mathrm{H}_{\min}(\Phi_1). \tag{8.442}$$

Define channels $\Psi_0, \Psi_1 \in \mathrm{C}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Y})$ as

$$\Psi_0(X) = \Phi_0(X) \otimes \Phi_1(\sigma_1) \quad \text{and} \quad \Psi_1(X) = \Phi_0(\sigma_0) \otimes \Phi_1(X) \tag{8.443}$$

for all $X \in \mathrm{L}(\mathcal{X})$, and define

$$\Phi = \Psi_0 \oplus \Psi_1 \in \mathrm{C}(\mathcal{X} \oplus \mathcal{X}, (\mathcal{Y} \otimes \mathcal{Y}) \oplus (\mathcal{Y} \otimes \mathcal{Y})). \tag{8.444}$$

It remains to verify that $\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi)$.

For any state $\rho \in \mathrm{D}(\mathcal{X} \oplus \mathcal{X})$, one may write

$$\rho = \begin{pmatrix} \lambda\rho_0 & X \\ X^* & (1-\lambda)\rho_1 \end{pmatrix} \tag{8.445}$$

for some choice of $\lambda \in [0,1]$, $\rho_0, \rho_1 \in \mathrm{D}(\mathcal{X})$, and $X \in \mathrm{L}(\mathcal{X})$. Evaluating Φ on ρ yields

$$\Phi(\rho) = \begin{pmatrix} \lambda\Phi_0(\rho_0) \otimes \Phi_1(\sigma_1) & 0 \\ 0 & (1-\lambda)\Phi_0(\sigma_0) \otimes \Phi_1(\rho_1) \end{pmatrix}, \tag{8.446}$$

so that

$$\begin{aligned} \mathrm{H}(\Phi(\rho)) &= \lambda\big(\mathrm{H}(\Phi_0(\rho_0)) + \mathrm{H}(\Phi_1(\sigma_1))\big) \\ &\quad + (1-\lambda)\big(\mathrm{H}(\Phi_0(\sigma_0)) + \mathrm{H}(\Phi_1(\rho_1))\big) + \mathrm{H}(\lambda, 1-\lambda). \end{aligned} \tag{8.447}$$

One concludes that

$$\mathrm{H}_{\min}(\Phi) = \mathrm{H}_{\min}(\Phi_0) + \mathrm{H}_{\min}(\Phi_1). \tag{8.448}$$

Finally, define an isometry $V \in \mathrm{U}(\mathcal{X} \otimes \mathcal{X}, (\mathcal{X} \oplus \mathcal{X}) \otimes (\mathcal{X} \oplus \mathcal{X}))$ by the equation

$$V(x_0 \otimes x_1) = (x_0 \oplus 0) \otimes (0 \oplus x_1) \tag{8.449}$$

holding for all $x_0, x_1 \in \mathcal{X}$. A calculation reveals that

$$\begin{aligned} &\mathrm{H}\big((\Phi \otimes \Phi)(V\xi V^*)\big) \\ &\quad = \mathrm{H}\big((\Phi_0 \otimes \Phi_1)(\xi)\big) + \mathrm{H}(\Phi_0(\sigma_0)) + \mathrm{H}(\Phi_1(\sigma_1)) \end{aligned} \tag{8.450}$$

for every density operator $\xi \in \mathrm{D}(\mathcal{X} \otimes \mathcal{X})$. In particular, for any choice of $\xi \in \mathrm{D}(\mathcal{X} \otimes \mathcal{X})$ satisfying

$$\mathrm{H}((\Phi_0 \otimes \Phi_1)(\xi)) = \mathrm{H}_{\min}(\Phi_0 \otimes \Phi_1), \tag{8.451}$$

one has

$$\begin{aligned} &\mathrm{H}\big((\Phi \otimes \Phi)(V\xi V^*)\big) \\ &\quad = \mathrm{H}_{\min}(\Phi_0 \otimes \Phi_1) + \mathrm{H}_{\min}(\Phi_0) + \mathrm{H}_{\min}(\Phi_1), \end{aligned} \tag{8.452}$$

and therefore $\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi)$ as claimed. □

From low minimum output entropy to high Holevo capacity

The construction to be described below allows one to conclude that there exists a channel Ψ for which the Holevo capacity is super-additive, meaning that

$$\chi(\Psi \otimes \Psi) > 2\chi(\Psi), \tag{8.453}$$

by means of Corollary 8.62.

Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are complex Euclidean spaces and $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ is an arbitrary channel. Suppose further that Σ is an alphabet and

$$\{U_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{Y}) \tag{8.454}$$

is a collection of unitary operators with the property that the completely depolarizing channel $\Omega \in \mathrm{C}(\mathcal{Y})$ is given by

$$\Omega(Y) = \frac{1}{|\Sigma|} \sum_{a \in \Sigma} U_a Y U_a^* \tag{8.455}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$. (Such a collection may, for instance, be derived from the discrete Weyl operators defined in Section 4.1.2.) Let $\mathcal{Z} = \mathbb{C}^{\Sigma}$ and define a new channel $\Psi \in \mathrm{C}(\mathcal{Z} \otimes \mathcal{X}, \mathcal{Y})$ by the equation

$$\Psi(E_{a,b} \otimes X) = \begin{cases} U_a \Phi(X) U_a^* & \text{if } a = b \\ 0 & \text{otherwise} \end{cases} \tag{8.456}$$

holding for all $a, b \in \Sigma$ and $X \in \mathrm{L}(\mathcal{X})$.

In more intuitive terms, the action of the channel Ψ may be described as follows. A pair of registers (Z, X) is taken as input, and a measurement of the register Z with respect to the standard basis of $\mathcal{Z}$ is made, yielding a symbol $a \in \Sigma$. The channel Φ is applied to X, resulting in a register Y, and the unitary channel described by U_a is applied to Y. The measurement outcome a is discarded and Y is taken to be the output of the channel.

As the following proposition shows, the Holevo capacity of the channel Ψ constructed in this way is determined by the minimum output entropy of the channel Φ.

Proposition 8.63. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. Also let Σ be an alphabet, let*

$$\{U_a : a \in \Sigma\} \subset \mathrm{U}(\mathcal{Y}) \tag{8.457}$$

be a collection of unitary operators for which the equation (8.455) *holds for all* $Y \in \mathrm{L}(\mathcal{Y})$*, let* $\mathcal{Z} = \mathbb{C}^{\Sigma}$*, and let* $\Psi \in \mathrm{C}(\mathcal{Z} \otimes \mathcal{X}, \mathcal{Y})$ *be a channel defined by the equation* (8.456) *holding for all* $a, b \in \Sigma$ *and* $X \in \mathrm{L}(\mathcal{X})$*. It holds that*

$$\chi(\Psi) = \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi). \tag{8.458}$$

Proof. Consider first the ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathcal{Z} \otimes \mathcal{X})$ defined as

$$\eta(a) = \frac{1}{|\Sigma|} E_{a,a} \otimes \rho \tag{8.459}$$

for all $a \in \Sigma$, where $\rho \in \mathrm{D}(\mathcal{X})$ is any state for which

$$\mathrm{H}_{\min}(\Phi) = \mathrm{H}(\Phi(\rho)). \tag{8.460}$$

One has

$$\begin{aligned} \chi(\Psi(\eta)) &= \mathrm{H}\left(\frac{1}{|\Sigma|} \sum_{a \in \Sigma} U_a \Phi(\rho) U_a^*\right) - \frac{1}{|\Sigma|} \sum_{a \in \Sigma} \mathrm{H}\big(U_a \Phi(\rho) U_a^*\big) \\ &= \mathrm{H}\big(\Omega(\rho)\big) - \mathrm{H}(\Phi(\rho)) \\ &= \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi). \end{aligned} \tag{8.461}$$

It therefore holds that

$$\chi(\Psi) \geq \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi). \tag{8.462}$$

Next, consider an arbitrary state $\sigma \in \mathrm{D}(\mathcal{Z} \otimes \mathcal{X})$. For $\Delta \in \mathrm{C}(\mathcal{Z})$ denoting the completely dephasing channel, one may write

$$\big(\Delta \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(\sigma) = \sum_{a \in \Sigma} q(a) E_{a,a} \otimes \xi_a, \tag{8.463}$$

for some choice of a probability vector $q \in \mathcal{P}(\Sigma)$ and a collection of states

$$\{\xi_a : a \in \Sigma\} \subseteq \mathrm{D}(\mathcal{X}). \tag{8.464}$$

It holds that

$$\Psi(\sigma) = \sum_{a \in \Sigma} q(a) U_a \Phi(\xi_a) U_a^* \tag{8.465}$$

and therefore

$$\mathrm{H}(\Psi(\sigma)) \geq \sum_{a \in \Sigma} q(a)\, \mathrm{H}\big(\Phi(\xi_a)\big) \geq \mathrm{H}_{\min}(\Phi) \tag{8.466}$$

by the concavity of the von Neumann entropy function (Theorem 5.25).

Finally, consider an arbitrary ensemble $\eta : \Gamma \to \mathrm{Pos}(\mathcal{Z} \otimes \mathcal{X})$, written as

$$\eta(b) = p(b)\sigma_b \tag{8.467}$$

for each $b \in \Gamma$, for $p \in \mathcal{P}(\Gamma)$ being a probability vector and

$$\{\sigma_b \,:\, b \in \Gamma\} \subseteq \mathrm{D}(\mathcal{Z} \otimes \mathcal{X}) \tag{8.468}$$

being a collection of states. It holds that

$$\begin{aligned} \chi(\Psi(\eta)) = \mathrm{H}\left(\sum_{b\in\Gamma} p(b)\Psi(\sigma_b)\right) - \sum_{b\in\Gamma} p(b)\,\mathrm{H}(\Psi(\sigma_b)) \\ \leq \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi). \end{aligned} \tag{8.469}$$

The ensemble η was chosen arbitrarily, and therefore

$$\chi(\Psi) \leq \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi), \tag{8.470}$$

which completes the proof. □

Theorem 8.64. *There exists a channel $\Psi \in \mathrm{C}(\mathcal{W}, \mathcal{Y})$, for some choice of complex Euclidean spaces $\mathcal{W}$ and $\mathcal{Y}$, such that*

$$\chi(\Psi \otimes \Psi) > 2\chi(\Psi). \tag{8.471}$$

Proof. By Corollary 8.62 there exist complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$ and a channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ for which the inequality

$$\mathrm{H}_{\min}(\Phi \otimes \Phi) < 2\,\mathrm{H}_{\min}(\Phi) \tag{8.472}$$

holds. Let Σ be an alphabet and let

$$\{U_a \,:\, a \in \Sigma\} \subset \mathrm{U}(\mathcal{Y}) \tag{8.473}$$

be a collection of unitary operators for which

$$\Omega(Y) = \frac{1}{|\Sigma|} \sum_{a\in\Sigma} U_a Y U_a^* \tag{8.474}$$

for all $Y \in \mathrm{L}(\mathcal{Y})$. Also let $\mathcal{Z} = \mathbb{C}^\Sigma$ and let $\Psi \in \mathrm{C}(\mathcal{Z} \otimes \mathcal{X}, \mathcal{Y})$ be a channel defined by the equation (8.456) above for all $a, b \in \Sigma$ and $X \in \mathrm{L}(\mathcal{X})$.

Up to a permutation of the tensor factors of its input space, $\Psi \otimes \Psi$ is equivalent to the channel $\Xi \in \mathrm{C}((\mathcal{Z} \otimes \mathcal{Z}) \otimes (\mathcal{X} \otimes \mathcal{X}), \mathcal{Y} \otimes \mathcal{Y})$ that would be obtained from the channel $\Phi \otimes \Phi$ through a similar construction, using the collection of unitary operators

$$\{U_a \otimes U_b \,:\, (a,b) \in \Sigma \times \Sigma\} \subset \mathrm{U}(\mathcal{Y} \otimes \mathcal{Y}). \tag{8.475}$$

It therefore follows from Proposition 8.63 that

$$\chi(\Psi) = \log(\dim(\mathcal{Y})) - \mathrm{H}_{\min}(\Phi) \tag{8.476}$$

while

$$\chi(\Psi \otimes \Psi) = \log(\dim(\mathcal{Y} \otimes \mathcal{Y})) - \mathrm{H}_{\min}(\Phi \otimes \Phi) > 2\chi(\Psi). \tag{8.477}$$

Taking $\mathcal{W} = \mathcal{Z} \otimes \mathcal{X}$, the theorem is therefore proved. □

One consequence of this theorem is that an analogous statement to the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30), but without a regularization, does not hold in general; because

$$\mathrm{C}(\Phi) \geq \frac{\chi(\Phi \otimes \Phi)}{2}, \tag{8.478}$$

it is the case that $\mathrm{C}(\Phi) > \chi(\Phi)$ for some choices of a channel Φ.

8.3.2 Super-activation of quantum channel capacity

The purpose of the present subsection is to demonstrate the phenomenon of *super-activation*, in which the tensor product of two zero-capacity channels have positive quantum capacity. As a byproduct, one obtains an example of a channel Ψ satisfying

$$\mathrm{I}_{\mathrm{c}}(\Psi \otimes \Psi) > 2\,\mathrm{I}_{\mathrm{c}}(\Psi). \tag{8.479}$$

Two classes of zero-capacity channels

It is possible to prove that certain classes of channels have zero quantum capacity. Channels whose Choi operators are PPT and self-complementary channels fall into this category. The following proposition establishes that channels whose Choi operators are PPT must have zero capacity.

Proposition 8.65. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel such that $J(\Phi) \in \mathrm{PPT}(\mathcal{Y} : \mathcal{X})$. It holds that $\mathrm{Q}(\Phi) = 0$.*

Proof. The first step of the proof is to establish that, for every choice of a complex Euclidean space $\mathcal{W}$ and a density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W})$, one has

$$\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\rho) \in \mathrm{PPT}(\mathcal{Y} : \mathcal{W}). \tag{8.480}$$

Toward this goal, observe that, for any choice of a complex Euclidean space $\mathcal{W}$ and a positive semidefinite operator $P \in \mathrm{Pos}(\mathcal{X} \otimes \mathcal{W})$, there must exist a completely positive map $\Psi_P \in \mathrm{CP}(\mathcal{X}, \mathcal{W})$ satisfying

$$P = \big(\mathbb{1}_{\mathrm{L}(\mathcal{X})} \otimes \Psi_P\big)\big(\mathrm{vec}(\mathbb{1}_{\mathcal{X}})\,\mathrm{vec}(\mathbb{1}_{\mathcal{X}})^*\big). \tag{8.481}$$

(The map Ψ_P is uniquely defined by this requirement—one may obtain its Choi representation by swapping the tensor factors of P.) It follows that, for any complex Euclidean space $\mathcal{W}$ and any density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W})$, one must have

$$\begin{aligned} &\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)\big(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})}\big)(\rho) \\ &\quad = \big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \Psi_\rho\big)\big(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{X})}\big)(J(\Phi)) \in \mathrm{Pos}(\mathcal{Y} : \mathcal{W}) \end{aligned} \tag{8.482}$$

by virtue of the fact that Ψ_ρ is completely positive and $J(\Phi) \in \mathrm{PPT}(\mathcal{Y} : \mathcal{X})$, which establishes (8.480).

Now, it follows from the assumption $J(\Phi) \in \mathrm{PPT}(\mathcal{Y} : \mathcal{X})$ that

$$J\big(\Phi^{\otimes n}\big) \in \mathrm{PPT}\big(\mathcal{Y}^{\otimes n} : \mathcal{X}^{\otimes n}\big) \tag{8.483}$$

for every positive integer n. For every choice of positive integers n and m, for $\mathcal{Z} = \mathbb{C}^\Gamma$ for $\Gamma = \{0, 1\}$, and for any channel $\Xi \in \mathrm{C}(\mathcal{Y}^{\otimes n}, \mathcal{Z}^{\otimes m})$, it therefore holds that

$$\big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho) \in \mathrm{PPT}\big(\mathcal{Z}^{\otimes m} : \mathcal{Z}^{\otimes m}\big) \tag{8.484}$$

for every density operator $\rho \in \mathrm{D}(\mathcal{X}^{\otimes n} \otimes \mathcal{Z}^{\otimes m})$. By Proposition 6.47, one therefore has

$$\mathrm{F}\Big(2^{-m}\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)\,\mathrm{vec}\big(\mathbb{1}_{\mathcal{Z}}^{\otimes m}\big)^*,\ \big(\Xi\Phi^{\otimes n} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})}^{\otimes m}\big)(\rho)\Big) \leq 2^{-m/2}. \tag{8.485}$$

For every choice of a positive real number $\alpha > 0$, it must therefore be the case that α fails to be an achievable rate for entanglement generation though Φ. Consequently, Φ has zero capacity for entanglement generation, which implies $\mathrm{Q}(\Phi) = 0$ by Theorem 8.50. □

The second category of channels mentioned above having zero quantum capacity are *self-complementary* channels. These are channels $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ such that there exists an isometry $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Y})$ such that

$$\Phi(X) = \big(\mathbb{1}_{\mathrm{L}(\mathcal{Y})} \otimes \mathrm{Tr}\big)(AXA^*) = \big(\mathrm{Tr} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}\big)(AXA^*) \tag{8.486}$$

for every $X \in \mathrm{L}(\mathcal{X})$.

It follows from Proposition 8.17 that the coherent information of every state $\sigma \in \mathrm{D}(\mathcal{X})$ through a self-complementary channel Φ must be zero:

$$\mathrm{I}_{\mathrm{C}}(\sigma; \Phi) = \mathrm{H}(\Phi(\sigma)) - \mathrm{H}(\Phi(\sigma)) = 0. \tag{8.487}$$

As every tensor power of a self-complementary channel is necessarily self-complementary, the quantum capacity theorem (Theorem 8.59) implies that self-complementary channels have zero quantum capacity. The following proposition states a more general variant of this observation.

Proposition 8.66. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be complementary channels, and suppose that there exists a channel $\Xi \in \mathrm{C}(\mathcal{Z}, \mathcal{Y})$ such that $\Phi = \Xi\Psi$. It holds that Φ has zero quantum capacity: $\mathrm{Q}(\Phi) = 0$.*

Proof. For every positive integer n, one has

$$\mathrm{I}_{\mathrm{C}}(\sigma; \Phi^{\otimes n}) = \mathrm{I}_{\mathrm{C}}(\sigma; \Xi^{\otimes n}\Psi^{\otimes n}) \leq \mathrm{I}_{\mathrm{C}}(\sigma; \Psi^{\otimes n}) \tag{8.488}$$

by Proposition 8.15. Because Ψ is complementary to Φ, it holds that $\Psi^{\otimes n}$ is complementary to $\Phi^{\otimes n}$ for every positive integer n, and therefore

$$\begin{aligned} \mathrm{I}_{\mathrm{C}}(\sigma; \Phi^{\otimes n}) &= \mathrm{H}\big(\Phi^{\otimes n}(\sigma)\big) - \mathrm{H}\big(\Psi^{\otimes n}(\sigma)\big) \\ &= -\mathrm{I}_{\mathrm{C}}(\sigma; \Psi^{\otimes n}) \leq -\mathrm{I}_{\mathrm{C}}(\sigma; \Phi^{\otimes n}), \end{aligned} \tag{8.489}$$

which implies $\mathrm{I}_{\mathrm{C}}(\sigma; \Phi^{\otimes n}) \leq 0$. By Theorem 8.59, it therefore holds that $\mathrm{Q}(\Phi) = 0$, which completes the proof. □

Remark 8.67. Channels of the form $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ for which there exists a channel $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ complementary to Φ, as well as a channel $\Xi \in \mathrm{C}(\mathcal{Z}, \mathcal{Y})$ for which $\Phi = \Xi\Psi$, are known as *anti-degradable channels*.

50% erasure channels

A *50%-erasure channel* is a simple type of self-complementary channel that plays a special role in the example of super-activation to be presented below. For any choice of a complex Euclidean space $\mathcal{X}$, the 50%-erasure channel defined with respect to $\mathcal{X}$ is the channel $\Xi \in \mathrm{C}(\mathcal{X}, \mathbb{C} \oplus \mathcal{X})$ defined as

$$\Xi(X) = \frac{1}{2} \begin{pmatrix} \mathrm{Tr}(X) & 0 \\ 0 & X \end{pmatrix} \tag{8.490}$$

for each $X \in \mathrm{L}(\mathcal{X})$.

Intuitively speaking, a 50%-erasure channel acts as the identity channel with probability $1/2$, and otherwise its input is lost (or erased). Under the assumption that $\mathcal{X} = \mathbb{C}^{\Sigma}$, for Σ being a given alphabet, one may associate the complex Euclidean space $\mathbb{C} \oplus \mathcal{X}$ with $\mathbb{C}^{\{\#\} \cup \Sigma}$, for # being a special *blank symbol* that is not contained in Σ. With this interpretation, the event that the input is erased may be associated with the blank symbol # being produced, so that

$$\Xi(X) = \frac{1}{2} X + \frac{1}{2} \mathrm{Tr}(X) E_{\#,\#} \tag{8.491}$$

for every $X \in \mathrm{L}(\mathcal{X})$. It should be noted that there is no ambiguity about whether an erasure has occurred for this channel. The situation is analogous to one in which a letter is sent through a postal system; and with probability $1/2$, the letter is received, and otherwise an empty envelope is received (as opposed to the letter being lost without the receiver's knowledge).

For every choice of $\mathcal{X}$, the 50%-erasure channel $\Xi \in \mathrm{C}(\mathcal{X}, \mathbb{C} \oplus \mathcal{X})$ is self-complementary: using the association of $\mathbb{C} \oplus \mathcal{X}$ with $\mathbb{C}^{\{\#\} \cup \Sigma}$, for Σ being a given alphabet that does not contain the blank symbol #, one has

$$\Xi(X) = \big(\mathrm{Tr} \otimes \mathbb{1}\big)\big(AXA^*\big) = \big(\mathbb{1} \otimes \mathrm{Tr}\big)\big(AXA^*\big) \tag{8.492}$$

for $A \in \mathrm{U}(\mathcal{X}, (\mathbb{C} \oplus \mathcal{X}) \otimes (\mathbb{C} \oplus \mathcal{X}))$ being the isometry defined as

$$A = \frac{1}{\sqrt{2}} \sum_{a \in \Sigma} (e_a \otimes e_{\#} + e_{\#} \otimes e_a) e_a^*. \tag{8.493}$$

It follows that $\mathrm{Q}(\Xi) = 0$.

A theorem of Smith and Yard

The following theorem allows one to prove lower bounds on the maximum coherent information of a channel tensored with a 50%-erasure channel on a sufficiently large space. For a suitable choice of a zero-capacity channel tensored with a 50%-erasure channel, the theorem leads to a demonstration of the super-activation phenomenon.

Theorem 8.68 (Smith–Yard)**.** *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be complex Euclidean spaces, let $A \in \mathrm{U}(\mathcal{X}, \mathcal{Y} \otimes \mathcal{Z})$ be an isometry, and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{X}, \mathcal{Z})$ be complementary channels defined as*

$$\Phi(X) = \mathrm{Tr}_{\mathcal{Z}}(AXA^*) \quad \text{and} \quad \Psi(X) = \mathrm{Tr}_{\mathcal{Y}}(AXA^*) \tag{8.494}$$

for every $X \in \mathrm{L}(\mathcal{X})$. Also let Σ be an alphabet, let $\eta : \Sigma \to \mathrm{Pos}(\mathcal{X})$ be an ensemble of states, and let $\mathcal{W}$ be a complex Euclidean space satisfying

$$\dim(\mathcal{W}) \geq \sum_{a \in \Sigma} \mathrm{rank}(\eta(a)). \tag{8.495}$$

There exists a density operator $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W})$ such that

$$\mathrm{I}_{\mathrm{c}}(\rho; \Phi \otimes \Xi) = \frac{1}{2}\chi(\Phi(\eta)) - \frac{1}{2}\chi(\Psi(\eta)), \tag{8.496}$$

for $\Xi \in \mathrm{C}(\mathcal{W}, \mathbb{C} \oplus \mathcal{W})$ denoting the 50%-erasure channel on $\mathcal{W}$.

Proof. By the assumption

$$\dim(\mathcal{W}) \geq \sum_{a \in \Sigma} \mathrm{rank}(\eta(a)), \tag{8.497}$$

one may choose a collection of vectors $\{u_a : a \in \Sigma\} \subset \mathcal{X} \otimes \mathcal{W}$ for which it holds that

$$\mathrm{Tr}_{\mathcal{W}}(u_a u_a^*) = \eta(a) \tag{8.498}$$

for each $a \in \Sigma$, and for which it holds that

$$\left\{\mathrm{Tr}_{\mathcal{X}}(u_a u_a^*) : a \in \Sigma\right\} \tag{8.499}$$

is an orthogonal set of operators. Let $\mathcal{V} = \mathbb{C}^{\Sigma}$, define a unit vector

$$u = \sum_{a \in \Sigma} e_a \otimes u_a \in \mathcal{V} \otimes \mathcal{X} \otimes \mathcal{W}, \tag{8.500}$$

and let $\rho = \mathrm{Tr}_{\mathcal{V}}(uu^*)$. One may observe that, by virtue of the fact that (8.499) is an orthogonal set, it holds that

$$\mathrm{Tr}_{\mathcal{W}}(uu^*) = \sum_{a\in\Sigma} E_{a,a} \otimes \eta(a). \tag{8.501}$$

For the unit vector $v \in \mathcal{V} \otimes \mathcal{Y} \otimes \mathcal{Z} \otimes \mathcal{W}$ defined as $v = (\mathbb{1}_{\mathcal{V}} \otimes A \otimes \mathbb{1}_{\mathcal{W}})u$, it therefore holds that

$$\mathrm{Tr}_{\mathcal{W}}(vv^*) = \sum_{a\in\Sigma} E_{a,a} \otimes A\eta(a)A^*. \tag{8.502}$$

The 50%-erasure channel Ξ has the property that

$$\mathrm{H}\big((\Phi\otimes\Xi)(\rho)\big) = \frac{1}{2}\mathrm{H}\big((\Phi\otimes\mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big) + \frac{1}{2}\mathrm{H}\big(\Phi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big) + 1, \tag{8.503}$$

and likewise for the channel Ψ in place of Φ. As Ψ is complementary to Φ and Ξ is self-complementary, it follows that

$$\begin{aligned}
\mathrm{I}_{\mathrm{C}}(\rho;\Phi\otimes\Xi) &= \mathrm{H}\big((\Phi\otimes\Xi)(\rho)\big) - \mathrm{H}\big((\Psi\otimes\Xi)(\rho)\big) \\
&= \frac{1}{2}\big(\mathrm{H}\big((\Phi\otimes\mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big) - \mathrm{H}\big((\Psi\otimes\mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big)\big) \\
&\quad + \frac{1}{2}\big(\mathrm{H}\big(\Phi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big) - \mathrm{H}\big(\Psi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big)\big).
\end{aligned} \tag{8.504}$$

Now, let V, Y, Z, and W be registers corresponding to the spaces $\mathcal{V}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$, respectively, and consider the situation in which the compound register $(\mathsf{V},\mathsf{Y},\mathsf{Z},\mathsf{W})$ is in the pure state vv^*. It holds that

$$\begin{aligned}
\mathrm{H}\big((\Phi\otimes\mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big) &= \mathrm{H}(\mathsf{Y},\mathsf{W}) = \mathrm{H}(\mathsf{V},\mathsf{Z}),\\
\mathrm{H}\big((\Psi\otimes\mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho)\big) &= \mathrm{H}(\mathsf{Z},\mathsf{W}) = \mathrm{H}(\mathsf{V},\mathsf{Y}),\\
\mathrm{H}\big(\Phi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big) &= \mathrm{H}(\mathsf{Y}),\\
\mathrm{H}\big(\Psi(\mathrm{Tr}_{\mathcal{W}}(\rho))\big) &= \mathrm{H}(\mathsf{Z}),
\end{aligned} \tag{8.505}$$

and therefore

$$\mathrm{I}_{\mathrm{C}}(\rho;\Phi\otimes\Xi) = \frac{1}{2}\mathrm{I}(\mathsf{V}:\mathsf{Y}) - \frac{1}{2}\mathrm{I}(\mathsf{V}:\mathsf{Z}) = \frac{1}{2}\chi(\Phi(\eta)) - \frac{1}{2}\chi(\Psi(\eta)), \tag{8.506}$$

as required. □

An explicit example of super-activation

An example of the super-activation phenomenon, based on Theorem 8.68, will now be described. The first step is to define a zero-capacity channel Φ as follows. Let

$$\alpha = \sqrt{\sqrt{2} - 1}, \quad \beta = \sqrt{1 - \frac{1}{\sqrt{2}}}, \quad \text{and} \quad \gamma = \sqrt{\frac{1}{\sqrt{2}} - \frac{1}{2}}, \tag{8.507}$$

define $A_1, \ldots, A_6 \in \mathrm{L}(\mathbb{C}^4)$ as

$$\begin{aligned}
A_1 &= \begin{pmatrix} 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 \\ \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \end{pmatrix}, & A_2 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ -\gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \end{pmatrix}, \\
A_3 &= \begin{pmatrix} \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & A_4 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
A_5 &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta \end{pmatrix}, & A_6 &= \begin{pmatrix} 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \end{pmatrix},
\end{aligned} \tag{8.508}$$

and define $\Phi \in \mathrm{C}(\mathbb{C}^4)$ as

$$\Phi(X) = \sum_{k=1}^{6} A_k X A_k^* \tag{8.509}$$

for every $X \in \mathrm{L}(\mathbb{C}^4)$.

The fact that Φ is a zero-capacity channel follows from the fact that the Choi representation of Φ is a PPT operator. One way to verify this claim is to check that

$$\bigl(\mathrm{T} \otimes \mathbb{1}_{\mathrm{L}(\mathbb{C}^4)}\bigr)(J(\Phi)) = J(\Theta) \tag{8.510}$$

for $\Theta \in \mathrm{C}(\mathbb{C}^4)$ being the channel defined as

$$\Theta(X) = \sum_{k=1}^{6} B_k X B_k^* \tag{8.511}$$

for every $X \in \mathrm{L}(\mathbb{C}^4)$, where $B_1, \ldots, B_6 \in \mathrm{L}(\mathbb{C}^4)$ are as follows:

$$B_1 = \begin{pmatrix} 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 \\ \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 \end{pmatrix}, \qquad B_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ \gamma & 0 & 0 & 0 \\ 0 & -\gamma & 0 & 0 \end{pmatrix},$$

$$B_3 = \begin{pmatrix} \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad B_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{8.512}$$

$$B_5 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta \end{pmatrix}, \qquad B_6 = \begin{pmatrix} 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \end{pmatrix}.$$

It therefore follows from Proposition 8.65 that Φ has zero quantum capacity.

A channel complementary to Φ is given by $\Psi \in \mathrm{C}(\mathbb{C}^4, \mathbb{C}^6)$ defined as

$$\Psi(X) = \sum_{k=1}^{4} C_k X C_k^* \tag{8.513}$$

for every $X \in \mathrm{L}(\mathbb{C}^4)$, where

$$C_1 = \begin{pmatrix} 0 & 0 & \alpha & 0 \\ 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \end{pmatrix}, \qquad C_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha \\ 0 & 0 & 0 & 0 \\ \beta & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$C_3 = \begin{pmatrix} \gamma & 0 & 0 & 0 \\ -\gamma & 0 & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \beta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad C_4 = \begin{pmatrix} 0 & \gamma & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta \\ 0 & 0 & \beta & 0 \end{pmatrix}. \tag{8.514}$$

Finally, define $\Sigma = \{0,1\}$, define density operators

$$\sigma_0 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \sigma_1 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{8.515}$$

and define an ensemble $\eta : \Sigma \to \mathrm{Pos}(\mathbb{C}^4)$ as $\eta(0) = \sigma_0/2$ and $\eta(1) = \sigma_1/2$. It holds that

$$\Phi(\sigma_0) = \begin{pmatrix} \frac{2-\sqrt{2}}{2} & 0 & 0 & 0 \\ 0 & \frac{2-\sqrt{2}}{2} & 0 & 0 \\ 0 & 0 & \frac{\sqrt{2}-1}{2} & 0 \\ 0 & 0 & 0 & \frac{\sqrt{2}-1}{2} \end{pmatrix} \tag{8.516}$$

and

$$\Phi(\sigma_1) = \begin{pmatrix} \frac{\sqrt{2}-1}{2} & 0 & 0 & 0 \\ 0 & \frac{\sqrt{2}-1}{2} & 0 & 0 \\ 0 & 0 & \frac{2-\sqrt{2}}{2} & 0 \\ 0 & 0 & 0 & \frac{2-\sqrt{2}}{2} \end{pmatrix}, \tag{8.517}$$

while

$$\Psi(\sigma_0) = \Psi(\sigma_1) = \begin{pmatrix} \frac{\sqrt{2}-1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{2}-1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{2-\sqrt{2}}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2-\sqrt{2}}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{2-\sqrt{2}}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2-\sqrt{2}}{4} \end{pmatrix}. \tag{8.518}$$

One therefore has that

$$\chi(\Phi(\eta)) = \mathrm{H}\Big(\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}\Big) - \mathrm{H}\Big(\tfrac{2-\sqrt{2}}{2}, \tfrac{2-\sqrt{2}}{2}, \tfrac{\sqrt{2}-1}{2}, \tfrac{\sqrt{2}-1}{2}\Big) > \tfrac{1}{50}, \tag{8.519}$$

while

$$\chi(\Psi(\eta)) = 0. \tag{8.520}$$

By Theorem 8.68, there must exist a density operator $\rho \in \mathrm{D}(\mathbb{C}^4 \otimes \mathbb{C}^4)$ such that

$$\mathrm{I_c}(\rho; \Phi \otimes \Xi) > \frac{1}{100}, \tag{8.521}$$

for $\Xi \in \mathrm{C}(\mathbb{C}^4, \mathbb{C} \oplus \mathbb{C}^4)$ being a 50%-erasure channel. One therefore has that $\mathrm{Q}(\Phi) = \mathrm{Q}(\Xi) = 0$, while $\mathrm{Q}(\Phi \otimes \Xi) > 0$.

The need for a regularization in the quantum capacity theorem

The super-activation example described above illustrates that the maximum coherent information is not additive; one has

$$\mathrm{I}_{\mathrm{c}}(\Phi \otimes \Xi) > \mathrm{I}_{\mathrm{c}}(\Phi) + \mathrm{I}_{\mathrm{c}}(\Xi) \tag{8.522}$$

for the channels Φ and Ξ specified in that example. As these channels are different, it does not follow immediately that a strict inequality of the form

$$\mathrm{I}_{\mathrm{c}}(\Psi^{\otimes n}) > n\,\mathrm{I}_{\mathrm{c}}(\Psi) \tag{8.523}$$

holds for any choice of a channel Ψ and a positive integer n. It is possible, however, to conclude that such an inequality does hold (for $n = 2$) using a direct sum construction along similar lines to the one used in the context of the Holevo capacity and minimum output entropy. The following three propositions that concern direct sums of channels will be used to reach this conclusion.

Proposition 8.69. *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, $\mathcal{Y}_1$, $\mathcal{Z}_0$, and $\mathcal{Z}_1$ be complex Euclidean spaces, and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$, $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$, $\Psi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Z}_0)$, and $\Psi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Z}_1)$ be channels such that Ψ_0 is complementary to Φ_0 and Ψ_1 is complementary to Φ_1. The channel $\Psi_0 \oplus \Psi_1$ is complementary to $\Phi_0 \oplus \Phi_1$.*

Proof. Let $A_0 \in \mathrm{U}(\mathcal{X}_0, \mathcal{Y}_0 \otimes \mathcal{Z}_0)$ and $A_1 \in \mathrm{U}(\mathcal{X}_1, \mathcal{Y}_1 \otimes \mathcal{Z}_1)$ be isometries such that the following equations hold for all $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$:

$$\begin{aligned} \Phi_0(X_0) &= \mathrm{Tr}_{\mathcal{Z}_0}\bigl(A_0 X_0 A_0^*\bigr), \qquad & \Psi_0(X_0) &= \mathrm{Tr}_{\mathcal{Y}_0}\bigl(A_0 X_0 A_0^*\bigr), \\ \Phi_1(X_1) &= \mathrm{Tr}_{\mathcal{Z}_1}\bigl(A_1 X_1 A_1^*\bigr), \qquad & \Psi_1(X_1) &= \mathrm{Tr}_{\mathcal{Y}_1}\bigl(A_1 X_1 A_1^*\bigr). \end{aligned} \tag{8.524}$$

Let $W \in \mathrm{U}\bigl((\mathcal{Y}_0 \otimes \mathcal{Z}_0) \oplus (\mathcal{Y}_1 \otimes \mathcal{Z}_1), (\mathcal{Y}_0 \oplus \mathcal{Y}_1) \otimes (\mathcal{Z}_0 \oplus \mathcal{Z}_1)\bigr)$ be the isometry defined by the equation

$$\begin{aligned} &W\bigl((y_0 \otimes z_0) \oplus (y_1 \otimes z_1)\bigr) \\ &\quad = (y_0 \oplus 0) \otimes (z_0 \oplus 0) + (0 \oplus y_1) \otimes (0 \oplus z_1) \end{aligned} \tag{8.525}$$

for every $y_0 \in \mathcal{Y}_0$, $y_1 \in \mathcal{Y}_1$, $z_0 \in \mathcal{Z}_0$, and $z_1 \in \mathcal{Z}_1$. The equations

$$\begin{aligned} (\Phi_0 \oplus \Phi_1)(X) &= \mathrm{Tr}_{\mathcal{Z}_0 \oplus \mathcal{Z}_1}\left(W \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix} X \begin{pmatrix} A_0^* & 0 \\ 0 & A_1^* \end{pmatrix} W^*\right) \\ (\Psi_0 \oplus \Psi_1)(X) &= \mathrm{Tr}_{\mathcal{Y}_0 \oplus \mathcal{Y}_1}\left(W \begin{pmatrix} A_0 & 0 \\ 0 & A_1 \end{pmatrix} X \begin{pmatrix} A_0^* & 0 \\ 0 & A_1^* \end{pmatrix} W^*\right) \end{aligned} \tag{8.526}$$

hold for all $X \in \mathrm{L}(\mathcal{X}_0 \oplus \mathcal{X}_1)$, which implies that $\Psi_0 \oplus \Psi_1$ is complementary to $\Phi_0 \oplus \Phi_1$, as required. □

Proposition 8.70. *Let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels, for $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ being complex Euclidean spaces, and let $\sigma \in \mathrm{D}(\mathcal{X}_0 \oplus \mathcal{X}_1)$ be an arbitrary density operator, written as*

$$\sigma = \begin{pmatrix} \lambda\sigma_0 & X \\ X^* & (1-\lambda)\sigma_1 \end{pmatrix} \tag{8.527}$$

for $\lambda \in [0,1]$, $\sigma_0 \in \mathrm{D}(\mathcal{X}_0)$, $\sigma_1 \in \mathrm{D}(\mathcal{X}_1)$, and $X \in \mathrm{L}(\mathcal{X}_1, \mathcal{X}_0)$. It holds that

$$\mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \oplus \Phi_1) = \lambda\, \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0) + (1-\lambda)\, \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1). \tag{8.528}$$

Proof. Observe first that

$$\begin{aligned} \mathrm{H}((\Phi_0 \oplus \Phi_1)(\sigma)) &= \mathrm{H} \begin{pmatrix} \lambda\Phi_0(\sigma_0) & 0 \\ 0 & (1-\lambda)\Phi_1(\sigma_1) \end{pmatrix} \\ &= \lambda\, \mathrm{H}(\Phi_0(\sigma_0)) + (1-\lambda)\, \mathrm{H}(\Phi_1(\sigma_1)) + \mathrm{H}(\lambda, 1-\lambda). \end{aligned} \tag{8.529}$$

Assuming that $\mathcal{Z}_0$ and $\mathcal{Z}_1$ are complex Euclidean spaces and $\Psi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Z}_0)$ and $\Psi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Z}_1)$ are channels complementary to Φ_0 and Φ_1, respectively, one has that

$$\begin{aligned} &\mathrm{H}((\Psi_0 \oplus \Psi_1)(\sigma)) \\ &\quad = \lambda\, \mathrm{H}(\Psi_0(\sigma_0)) + (1-\lambda)\, \mathrm{H}(\Psi_1(\sigma_1)) + \mathrm{H}(\lambda, 1-\lambda) \end{aligned} \tag{8.530}$$

by a similar calculation to (8.529). As $\Psi_0 \oplus \Psi_1$ is complementary to $\Phi_0 \oplus \Phi_1$, as established in Proposition 8.69, it follows that

$$\begin{aligned} \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \oplus \Phi_1) &= \mathrm{H}((\Phi_0 \oplus \Phi_1)(\sigma)) - \mathrm{H}((\Psi_0 \oplus \Psi_1)(\sigma)) \\ &= \lambda\big(\mathrm{H}(\Phi_0(\sigma_0)) - \mathrm{H}(\Psi_0(\sigma_0))\big) \\ &\qquad + (1-\lambda)\big(\mathrm{H}(\Phi_1(\sigma_1)) - \mathrm{H}(\Psi_1(\sigma_1))\big) \\ &= \lambda\, \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi_0) + (1-\lambda)\, \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi_1) \end{aligned} \tag{8.531}$$

as required. □

Proposition 8.71. *Let $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$ be complex Euclidean spaces and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels. It holds that*

$$\mathrm{I}_{\mathrm{c}}((\Phi_0 \oplus \Phi_1) \otimes (\Phi_0 \oplus \Phi_1)) \geq \mathrm{I}_{\mathrm{c}}(\Phi_0 \otimes \Phi_1). \tag{8.532}$$

Proof. Define an isometry $W \in \mathrm{U}\big(\mathcal{X}_0 \otimes \mathcal{X}_1, (\mathcal{X}_0 \oplus \mathcal{X}_1) \otimes (\mathcal{X}_0 \oplus \mathcal{X}_1)\big)$ by the equation

$$W(x_0 \otimes x_1) = (x_0 \oplus 0) \otimes (0 \oplus x_1) \tag{8.533}$$

holding for all $x_0 \in \mathcal{X}_0$ and $x_1 \in \mathcal{X}_1$, and along similar lines, define an isometry $V \in \mathrm{U}\big(\mathcal{Y}_0 \otimes \mathcal{Y}_1, (\mathcal{Y}_0 \oplus \mathcal{Y}_1) \otimes (\mathcal{Y}_0 \oplus \mathcal{Y}_1)\big)$ by the equation

$$V(y_0 \otimes y_1) = (y_0 \oplus 0) \otimes (0 \oplus y_1) \tag{8.534}$$

for all $y_0 \in \mathcal{Y}_0$ and $y_1 \in \mathcal{Y}_1$. One has that

$$\begin{aligned} \big((\Phi_0 \oplus \Phi_1) \otimes (\Phi_0 \oplus \Phi_1)\big)(W(X_0 \otimes X_1)W^*) & \\ = \begin{pmatrix} \Phi_0(X_0) & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & \Phi_1(X_1) \end{pmatrix} & \\ = V(\Phi_0(X_0) \otimes \Phi_1(X_1))V^* & \end{aligned} \tag{8.535}$$

for all $X_0 \in \mathrm{L}(\mathcal{X}_0)$ and $X_1 \in \mathrm{L}(\mathcal{X}_1)$. For every choice of a density operator $\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)$, it follows that

$$\mathrm{I}_{\mathrm{c}}(W\sigma W^*; (\Phi_0 \oplus \Phi_1) \otimes (\Phi_0 \oplus \Phi_1)) = \mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1), \tag{8.536}$$

which implies the proposition. □

Finally, consider the channel $\Psi = \Phi \oplus \Xi$, for Φ and Ξ as in the example of super-activation described above. By Proposition 8.70, one may conclude that $\mathrm{I}_{\mathrm{c}}(\Phi \oplus \Xi) = 0$, while Proposition 8.71 implies

$$\mathrm{I}_{\mathrm{c}}\big((\Phi \oplus \Xi) \otimes (\Phi \oplus \Xi)\big) \geq \mathrm{I}_{\mathrm{c}}(\Phi \otimes \Xi) > 0. \tag{8.537}$$

It therefore holds that the channel $\Psi = \Phi \oplus \Xi$ satisfies the strict inequality (8.523) for $n = 2$.

As a consequence of this fact, one has that the quantum capacity and maximum coherent information differ for some channels. In this sense, the regularization in the quantum capacity theorem (Theorem 8.59) is similar to the one in the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30) in that it cannot generally be removed.

8.4 Exercises

8.1. Let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be channels, for an arbitrary choice of complex Euclidean spaces $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{Y}_0$, and $\mathcal{Y}_1$.

(a) Prove that

$$\mathrm{I}_{\mathrm{c}}(\Phi_0 \oplus \Phi_1) = \max\bigl\{\mathrm{I}_{\mathrm{c}}(\Phi_0), \mathrm{I}_{\mathrm{c}}(\Phi_1)\bigr\}. \tag{8.538}$$

(b) Prove that

$$\chi(\Phi_0 \oplus \Phi_1) = \max_{\lambda \in [0,1]} \Bigl(\lambda \chi(\Phi_0) + (1-\lambda)\chi(\Phi_1) + \mathrm{H}(\lambda, 1-\lambda)\Bigr). \tag{8.539}$$

8.2. Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be complex Euclidean spaces, let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ and $\Psi \in \mathrm{C}(\mathcal{Z}, \mathcal{W})$ be channels, and assume that Φ is an entanglement breaking channel (q.v. Exercise 6.1). Prove that the following identities hold:

(a) $\mathrm{H}_{\min}(\Phi \otimes \Psi) = \mathrm{H}_{\min}(\Phi) + \mathrm{H}_{\min}(\Psi)$.

(b) $\chi(\Phi \otimes \Psi) = \chi(\Phi) + \chi(\Psi)$.

(c) $\mathrm{I}_{\mathrm{c}}(\Phi \otimes \Psi) = \mathrm{I}_{\mathrm{c}}(\Psi)$.

The inequality established by a correct answer to Exercise 5.6 is useful for solving this exercise.

8.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be complex Euclidean spaces and let $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$ be a channel. It is said that Φ is *degradable* if and only if there exists a complex Euclidean space $\mathcal{Z}$ and a channel $\Psi \in \mathrm{C}(\mathcal{Y}, \mathcal{Z})$ for which it holds that $\Psi\Phi$ is complementary to Φ.

(a) Prove that, for any choice of a degradable channel $\Phi \in \mathrm{C}(\mathcal{X}, \mathcal{Y})$, states $\sigma_0, \sigma_1 \in \mathrm{D}(\mathcal{X})$, and a real number $\lambda \in [0,1]$, the following inequality holds:

$$\mathrm{I}_{\mathrm{c}}\bigl(\lambda\sigma_0 + (1-\lambda)\sigma_1; \Phi\bigr) \geq \lambda\, \mathrm{I}_{\mathrm{c}}(\sigma_0; \Phi) + (1-\lambda)\, \mathrm{I}_{\mathrm{c}}(\sigma_1; \Phi). \tag{8.540}$$

(Equivalently, the function $\sigma \mapsto \mathrm{I}_{\mathrm{c}}(\sigma; \Phi)$ defined on $\mathrm{D}(\mathcal{X})$ is concave.)

(b) Let X_0, X_1, Y_0, and Y_1 be registers, let $\sigma \in \mathrm{D}(\mathcal{X}_0 \otimes \mathcal{X}_1)$ be an arbitrary state of the pair $(\mathsf{X}_0, \mathsf{X}_1)$, and let $\Phi_0 \in \mathrm{C}(\mathcal{X}_0, \mathcal{Y}_0)$ and $\Phi_1 \in \mathrm{C}(\mathcal{X}_1, \mathcal{Y}_1)$ be degradable channels. Prove that

$$\mathrm{I}_{\mathrm{c}}(\sigma; \Phi_0 \otimes \Phi_1) = \mathrm{I}_{\mathrm{c}}(\sigma[\mathsf{X}_0]; \Phi_0) + \mathrm{I}_{\mathrm{c}}(\sigma[\mathsf{X}_1]; \Phi_1). \tag{8.541}$$

8.4. Let $\mathcal{X}$ be a complex Euclidean space, let $\lambda \in [0,1]$, and define a channel $\Xi \in \mathrm{C}(\mathcal{X}, \mathbb{C} \oplus \mathcal{X})$ as

$$\Xi(X) = \begin{pmatrix} \lambda \operatorname{Tr}(X) & 0 \\ 0 & (1-\lambda)X \end{pmatrix} \tag{8.542}$$

for all $X \in \mathrm{L}(\mathcal{X})$.

(a) Give a closed-form expression for the coherent information $\mathrm{I}_{\mathrm{C}}(\sigma; \Xi)$ of an arbitrary state $\sigma \in \mathrm{D}(\mathcal{X})$ through Ξ.

(b) Give a closed-form expression for the entanglement-assisted classical capacity $\mathrm{C}_{\mathrm{E}}(\Xi)$ of Ξ.

(c) Give a closed-form expression for the quantum capacity $\mathrm{Q}(\Xi)$ of Ξ.

The results of Exercise 8.3 may be helpful when solving this exercise. The closed-form expressions should be functions of λ and $n = \dim(\mathcal{X})$ alone.

8.5. Let n be a positive integer, let $\mathcal{X} = \mathbb{C}^{\mathbb{Z}_n}$, and let $\{W_{a,b} : a, b \in \mathbb{Z}_n\}$ denote the set of discrete Weyl operators acting on $\mathcal{X}$ (q.v. Section 4.1.2 of Chapter 4). Also let $p \in \mathcal{P}(\mathbb{Z}_n)$ be a probability vector, and define a channel $\Phi \in \mathrm{C}(\mathcal{X})$ as

$$\Phi(X) = \sum_{a \in \mathbb{Z}_n} p(a) W_{0,a} X W_{0,a}^* \tag{8.543}$$

for all $X \in \mathrm{L}(\mathcal{X})$. Prove that

$$\mathrm{I}_{\mathrm{C}}(\Phi) = \log(n) - \mathrm{H}(p). \tag{8.544}$$

Like the previous problem, the results of Exercise 8.3 may be helpful when solving this exercise.

8.6. For every positive integer n and every real number $\varepsilon \in [0,1]$, define a channel $\Phi_{n,\varepsilon} \in \mathrm{C}(\mathbb{C}^n)$ as

$$\Phi_{n,\varepsilon} = \varepsilon \mathbb{1}_n + (1-\varepsilon)\Omega_n, \tag{8.545}$$

where $\mathbb{1}_n \in \mathrm{C}(\mathbb{C}^n)$ and $\Omega_n \in \mathrm{C}(\mathbb{C}^n)$ denote the identity and completely depolarizing channels defined with respect to the space $\mathbb{C}^n$. Prove that, for every choice of a positive real number K, there exists a choice of n and ε for which

$$\mathrm{C}_{\mathrm{E}}(\Phi_{n,\varepsilon}) \geq K \, \mathrm{C}(\Phi_{n,\varepsilon}) > 0. \tag{8.546}$$

8.5 Bibliographic remarks

The study of quantum channel capacities is, perhaps obviously, motivated in large part by Shannon's channel coding theorem [191], and the goal of obtaining analogous statements for quantum channels. It was soon realized, however, that there would not be a single capacity of a quantum channel, but rather several inequivalent but nevertheless fundamentally interesting capacities. The 1998 survey of Bennett and Shor [38] provides a summary of what was known about channel capacities at a relatively early point in their study.

Holevo [111] and Schumacher and Westmoreland [188] independently proved the Holevo–Schumacher–Westmoreland theorem (Theorem 8.30), in both cases building on Hausladen, Jozsa, Schumacher, Westmoreland, and Wootters [92]. The definition of what is now called the Holevo capacity (or the *Holevo information* of a channel) originates with the work of Holevo and Schumacher and Westmoreland. Lemma 8.28 was proved by Hayashi and Nagaoka [94], who used it in the analysis of generalizations of the Holevo–Schumacher–Westmoreland theorem.

The entanglement-assisted classical capacity theorem (Theorem 8.44) was proved by Bennett, Shor, Smolin, and Thapliyal [39]. The proof of this theorem presented in this chapter is due to Holevo [112]. Lemma 8.41 is due to Adami and Cerf [2].

Tasks involving quantum information transmission through quantum channels, along with fundamental definitions connected with such tasks, were investigated in several papers, including papers of Schumacher [186], Schumacher and Nielsen [187], Adami and Cerf [2], and Barnum, Nielsen, and Schumacher [22]. The entanglement generation capacity of a channel was defined explicitly by Devetak [58], and Theorems 8.49 and 8.50 follow from results proved by Barnum, Knill, and Nielsen [21].

The coherent information of a state through a channel was defined by Schumacher and Nielsen [187]. Lloyd [150] recognized the fundamental connection between the maximum coherent information of a channel and its quantum capacity, and provided a heuristic argument in support of the quantum capacity theorem (Theorem 8.59). The first rigorous proof of the quantum capacity theorem to be published was due to Devetak [58]. Shor reported a different proof of this theorem prior to Devetak's proof, although it was not published. A proof appearing in a subsequent paper of Hayden,

Shor, and Winter [98] reportedly resembles Shor's original proof.

The proof of the quantum capacity theorem presented in this chapter is due to Hayden, M. Horodecki, Winter, and Yard [95], incorporating some simplifying ideas due to Klesse [132], who independently proved the same theorem based on similar techniques. The phenomenon of decoupling (as represented by Lemma 8.53) provides a key step in this proof; this basic technique was used by Devetak [58], and was identified more explicitly by M. Horodecki, Oppenheim, and Winter [118] and Abeyesinghe, Devetak, Hayden, and Winter [1]. The PhD thesis of Dupuis [62] may be consulted for further information on this technique.

Shor [192] proved that the non-additivity of Holevo capacity follows from the non-additivity of minimum output entropy. In the same paper, Shor also proved the converse implication, which naturally had greater relevance prior to Hastings proof that the minimum output entropy is non-additive, along with the equivalence of these two non-additivity statements with two other statements concerning the entanglement of formation. The direct sum construction of channels and its implications to the additivity of channel capacities was investigated by Fukuda and Wolf [76].

DiVincenzo, Shor, and Smolin [61] proved that the coherent information is non-additive in 1998. Various properties of quantum erasure channels were established by Bennett, DiVincenzo, and Smolin [35]. Theorem 8.68, along with the realization that it gives an example of the super-activation phenomenon, is due to Smith and Yard [194]. The channel Φ described in the chapter giving rise to an example of super-activation, which appears in Smith and Yard's paper as well, was identified by K. Horodecki, Pankowski, M. Horodecki, and P. Horodecki [115], as it relates to a different capacity known as the *private capacity* of a channel.

Index

Bibliography

[1] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: restructuring quantum informationâĂŹs family tree. *Proceedings of the Royal Society A*, 465(2108):2537–2563, 2009.

[2] C. Adami and N. Cerf. Von Neumann capacity of noisy quantum channels. *Physical Review A*, 56(5):3470–3483, 1997.

[3] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 20–30, 1998.

[4] G. Alber, T. Beth, C. Charnes, A. Delgado, M. Grassl, and M. Mussinger. Stabilizing distinguishable qubits against spontaneous decay by detected-jump correcting quantum codes. *Physical Review Letters*, 86(19):4402–4405, 2001.

[5] P. Alberti. A note on the transition probability over C*-algebras. *Letters in Mathematical Physics*, 7(1):25–32, 1983.

[6] P. Alberti and A. Uhlmann. *Stochasticity and Partial Order*, volume 9 of *Mathematics and Its Applications*. D. Reidel Publishing Company, 1982.

[7] P. Alberti and A. Uhlmann. Stochastic linear maps and transition probability. *Letters in Mathematical Physics*, 7(2):107–112, 1983.

[8] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pages 376–383, 1999.

[9] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM*, 49(4):496–511, 2002.

[10] T. Ando. Concavity of certain maps on positive definite matrices and applications to Haramard products. *Linear Algebra and Its Applications*, 26:203–241, 1979.

[11] T. Apostol. *Mathematical Analysis*. Addison–Wesley, second edition, 1974.

[12] H. Araki and E. Lieb. Entropy inequalities. *Communications in Mathematical Physics*, 18(2):160–170, 1970.

[13] A. Arias, A. Gheondea, and S. Gudder. Fixed points of quantum operations. *Journal of Mathematical Physics*, 43(12):5872–5881, 2002.

[14] W. Arveson. Subalgebras of C*-algebras. *Acta Mathematica*, 123(1):141–224, 1969.

[15] R. Ash. *Information Theory*. Dover, 1990. Originally published in 1965 by Interscience Publishers.

[16] G. Aubrun, S. Szarek, and E. Werner. Hasting's additivity counterexample via Dvoretzky's theorem. *Communications in Mathematical Physics*, 305(1):85–97, 2011.

[17] K. Audenaert. A sharp Fannes-type inequality for the von Neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127–8136, 2007.

[18] K. Audenaert and S. Scheel. On random unitary channels. *New Journal of Physics*, 10:023011, 2008.

[19] S. Axler. *Linear Algebra Done Right*. Springer, second edition, 1997.

[20] H. Barnum and E. Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, 43(5):2097–2106, 2002.

[21] H. Barnum, E. Knill, and M. Nielsen. On quantum fidelities and channel capacities. *IEEE Transactions on Information Theory*, 46(4):1317–1329, 2000.

[22] H. Barnum, M. Nielsen, and B. Schumacher. Information transmission through a noisy quantum channel. *Physical Reivew A*, 57(6):4153–4175, 1998.

[23] J. Barrett. Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a Bell inequality. *Physical Review A*, 65(4):042302, 2002.

[24] R. Bartle. *The Elements of Integration*. John Wiley & Sons, Inc., 1966.

[25] D. Beckman, D. Gottesman, M. Nielsen, and J. Preskill. Causal and localizable quantum operations. *Physical Review A*, 64(5):52309, 2001.

[26] V. Belavkin. Optimal multiple quantum statistical hypothesis testing. *Stochastics*, 1:315–345, 1975.

[27] J. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1(3):195–200, 1964.

[28] A. Ben-Aroya and A. Ta-Shma. On the complexity of approximating the diamond norm. *Quantum Information and Computation*, 10(1):77–86, 2010.

[29] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States*. Cambridge University Press, 2006.

[30] C. Bennett, H. Bernstein, S. Popescu, and B. Schumacher. Concentrating partial entanglement by local operations. *Physical Review A*, 53(4):2046–2052, 1996.

[31] C. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. Wootters. Teleporting an unknown quantum state via dual classical and EPR channels. *Physical Review Letters*, 70(12):1895–1899, 1993.

[32] C. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. Smolin, and W. Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical Review Letters*, 76(5):722–725, 1996.

[33] C. Bennett, D. DiVincenzo, C. Fuchs, T. Mor, E. Rains, P. Shor, J. Smolin, and W. Wootters. Quantum nonlocality without entanglement. *Physical Review A*, 59:1070–1091, 1999.

[34] C. Bennett, D. DiVincenzo, T. Mor, P. Shor, J. Smolin, and B. Terhal. Unextendible product bases and bound entanglement. *Physical Review Letters*, 82(26):5385–5388, 1999.

[35] C. Bennett, D. DiVincenzo, and J. Smolin. Capacities of quantum erasure channels. *Physical Review Letters*, 78(16):3217–3220, 1997.

[36] C. Bennett, D. DiVincenzo, J. Smolin, and W. Wootters. Mixed-state entanglement and quantum error correction. *Physical Reivew A*, 54(5):3824–3851, 1996.

[37] C. Bennett, P. Hayden, D. Leung, P. Shor, and A. Winter. Remote preparation of quantum states. *IEEE Transactions on Information Theory*, 51(1):56–74, 2005.

[38] C. Bennett and P. Shor. Quantum information theory. *IEEE Transactions on Information Theory*, 44(6):2724–2742, 1998.

[39] C. Bennett, P. Shor, J. Smolin, and A. Thapliyal. Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters*, 83(15):3081–3084, 1999.

[40] C. Bennett and S. Wiesner. Communication via one-and two-particle operators on Einstein–Podolsky–Rosen states. *Physical Review Letters*, 69(20):2881–2884, 1992.

[41] R. Bhatia. *Matrix Analysis*. Springer, 1997.

[42] O. Bratteli, P. Jorgensen, A. Kishimoto, and R. Werner. Pure states on $\mathcal{O}_d$. *Journal of Operator Theory*, 43(1):97–143, 2000.

[43] F. Buscemi. On the minimum number of unitaries needed to describe a random-unitary channel. *Physics Letters A*, 360(2):256–258, 2006.

[44] C. Caves, C. Fuchs, and R. Schack. Unknown quantum states: the quantum de Finetti representation. *Journal of Mathematical Physics*, 43(9):4537–4559, 2002.

[45] A. Childs, D. Leung, L. Mančinska, and M. Ozols. A framework for bounding nonlocality of state discrimination. *Communications in Mathematical Physics*, 323(3):1121–1153, 2013.

[46] A. Childs, J. Preskill, and J. Renes. Quantum information and precision measurement. *Journal of Modern Optics*, 47(2–3):155–176, 2000.

[47] G. Chiribella, G. D'Ariano, and P. Perinotti. Transforming quantum operations: Quantum supermaps. *Europhysics Letters*, 83(3):30004, 2008.

[48] G. Chiribella, G. D'Ariano, and P. Perinotti. Theoretical framework for quantum networks. *Physical Review A*, 80(2):022339, 2009.

[49] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter. Everything you always wanted to know about LOCC (but were afraid to ask). *Communications in Mathematical Physics*, 328(1):303–326, 2014.

[50] M.-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and Its Applications*, 10(3):285–290, 1975.

[51] M. Christandl, R. König, G. Mitchison, and R. Renner. One-and-a-half quantum de Finetti theorems. *Communications in Mathematical Physics*, 273(2):473–498, 2007.

[52] J. Clauser, M. Horne, A. Shimony, and R. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969.

[53] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley Interscience, second edition, 2006.

[54] E. Davies. On the repeated measurement of continuous observables in quantum mechanics. *Journal of Functional Analysis*, 6(2):318–346, 1970.

[55] E. Davies and J. Lewis. An operational approach to quantum probability. *Communications in Mathematical Physics*, 17:239–260, 1970.

[56] B. de Finetti. La prévision : ses lois logiques, ses sources subjectives. *Annales de l'institut Henri Poincaré*, 7(1):1–68, 1937.

[57] J. de Pillis. Linear transformations which preserve Hermitian and positive semidefinite operators. *Pacific Journal of Mathematics*, 23(1):129–137, 1967.

[58] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51(1):44–55, 2005.

[59] P. Diaconis and D. Freedman. Finite exchangeable sequences. *The Annals of Probability*, 8(4):745–764, 1980.

[60] P. Diaconis and M. Shahshahani. The subgroup algorithm for generating uniform random variables. *Probability in the Engineering and Informational Sciences*, 1(1):15–32, 1987.

[61] D. DiVincenzo, P. Shor, and J. Smolin. Quantum-channel capacity of very noisy channels. *Physical Review A*, 57(2):830–839, 1998.

[62] F. Dupuis. *The Decoupling Approach to Quantum Information Theory*. PhD thesis, Univversité de Montréal, 2009.

[63] A. Dvoretzky. Some results on convex bodies and Banach spaces. In *Proceedings of the International Symposium on Linear Spaces (Held at the Hebrew University of Jerusalem, July 1960)*, pages 123–160, 1961.

[64] F. Dyson. Statistical theory of the energy levels of complex systems. I. *Journal of Mathematical Physics*, 3(1):140–156, 1962.

[65] F. Dyson. Statistical theory of the energy levels of complex systems. II. *Journal of Mathematical Physics*, 3(1):157–165, 1962.

[66] F. Dyson. Statistical theory of the energy levels of complex systems. III. *Journal of Mathematical Physics*, 3(1):166–175, 1962.

[67] T. Eggeling, D. Schlingemann, and R. Werner. Semicausal operations are semilocalizable. *Europhysics Letters*, 57(6):782–788, 2002.

[68] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10):777–780, 1935.

[69] Y. Eldar and D. Forney. On quantum detection and the square-root measurement. *IEEE Transactions on Information Theory*, 47(3):858–872, 2001.

[70] Y. Eldar, A. Megretski, and G. Verghese. Designing optimal quantum detectors via semidefinite programming. *IEEE Transactions on Information Theory*, 49(4):1007–1012, 2003.

[71] M. Fannes. A continuity property of the entropy density for spin lattie systems. *Communications in Mathematical Physics*, 31(4):291–294, 1973.

[72] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley & Sons, Inc., third edition, 1968.

[73] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume II. John Wiley & Sons, Inc., second edition, 1971.

[74] C. Fuchs and C. Caves. Mathematical techniques for quantum communication theory. *Open Systems & Information Dynamics*, 3(3):345–356, 1995.

[75] C. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999.

[76] M. Fukuda and M. Wolf. Simplifying additivity problems using direct sum constructions. *Journal of Mathematical Physics*, 48(7):072101, 2007.

[77] V. Gheorghiu and R. Griffiths. Separable operations of pure states. *Physical Review A*, 78(2):020304, 2008.

[78] A. Gilchrist, N. Langford, and M. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6):062310, 2005.

[79] R. Goodman and N. Wallach. *Representations and Invariants of the Classical Groups*, volume 68 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 1998.

[80] M. Gregoratti and R. Werner. Quantum lost and found. *Journal of Modern Optics*, 50(67):915–933, 2003.

[81] W. Greub. *Mutilinear Algebra*. Springer–Verlag, second edition, 1978.

[82] L. Gurvits. Classical deterministic complexity of Edmonds' problem and quantum entanglement. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pages 1–19, 2003.

[83] L. Gurvits and H. Barnum. Largest separable balls around the maximally mixed bipartite quantum state. *Physical Review A*, 66(6):062311, 2002.

[84] G. Gutoski and J. Watrous. Quantum interactive proofs with competing provers. In *Proceedings of the 22nd Symposium on Theoretical Aspects of Computer Science*, volume 3404 of *Lecture Notes in Computer Science*, pages 605–616. Springer, 2005.

[85] G. Gutoski and J. Watrous. Toward a general theory of quantum games. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 565–574, 2007.

[86] R. Haag and D. Kastler. An algebraic approach to quantum field theory. *Journal of Mathematical Physics*, 5(7):848–861, 1964.

[87] A. Haar. Der massbegriff in der theorie der kontinuierlichen gruppen. *Annals of Mathematics (Second Series)*, 34(1):147–169, 1933.

[88] P. Halmos. *Measure Theory*. Springer-Verlag, 1974. Originally published in 1950 by Litton Educational Publishing, Inc.

[89] P. Halmos. *Finite-Dimensional Vector Spaces*. Springer–Verlag, 1978. Originally published in 1942 by Princeton University Press.

[90] A. Harrow, P. Hayden, and D. Leung. Superdense coding of quantum states. *Physical Review Letters*, 92(18):187901, 2004.

[91] M. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5(4):255–257, 2009.

[92] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. Wootters. Classical information capacity of a quantum channel. *Physical Review A*, 54(3):1869–1876, 1996.

[93] P. Hausladen and W. Wootters. A 'pretty good' measurement for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, 1994.

[94] M. Hayashi and H. Nagaoka. General formulas for capacity of classical-quantum channels. *IEEE Transactions on Information Theory*, 49(7):1753–1768, 2003.

[95] P. Hayden, M. Horodecki, A. Winter, and J. Yard. A decoupling approach to the quantum capacity. *Open Systems & Information Dynamics*, 15(1):7–19, 2008.

[96] P. Hayden, D. Leung, P. Shor, and A.Winter. Randomizing quantum states: constructions and applications. *Communications in Mathematical Physics*, 250(2):371–391, 2004.

[97] P. Hayden, D. Leung, and A. Winter. Aspects of generic entanglement. *Communications in Mathematical Physics*, 265(1):95–117, 2006.

[98] P. Hayden, P. Shor, and A. Winter. Random quantum codes from Gaussian ensembles and an uncertainty relation. *Open Systems & Information Dynamics*, 15(1):71–89, 2008.

[99] P. Hayden and A. Winter. Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$. *Communications in Mathematical Physics*, 284(1):263–280, 2008.

[100] C. Helstrom. Detection theory and quantum mechanics. *Information and Control*, 10:254–291, 1967.

[101] C. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, 1976.

[102] F. Hiai, M. Ohya, and M. Tsukada. Sufficiency, KMS condition and relative entropy in von Neumann algebras. *Pacific Journal of Mathematics*, 96(1):99–109, 1981.

[103] K. Hoffman and R. Kunze. *Linear Algebra*. Prentice-Hall, Inc., second edition, 1971.

[104] A. Holevo. An analogue of statistical decision theory and noncommutative probability theory. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 26:133–149, 1972.

[105] A. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973.

[106] A. Holevo. Information-theoretical aspects of quantum measurement. *Problemy Peredachi Informatsii*, 9(2):31–42, 1973.

[107] A. Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3:337–394, 1973.

[108] A. Holevo. Statistical problems in quantum physics. In *Proceedings of the Second Japan-USSR Symposium on Probability Theory*, volume 330 of *Lecture Notes in Mathematics*, pages 104–119. Springer, 1973.

[109] A. Holevo. A note on covariant dynamical semigroups. *Reports on Mathematical Physics*, 32(2):211–216, 1993.

[110] A. Holevo. Covariant quantum Markovian evolutions. *Journal of Mathematical Physics*, 37(4):1812–1832, 1996.

[111] A. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998.

[112] A. Holevo. On entanglement-assisted classical capacity. *Journal of Mathematical Physics*, 43(9):4326–4333, 2002.

[113] A. Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. *American Journal of Mathematics*, 76(3):620–630, 1954.

[114] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[115] K. Horodecki, Ł. Pankowski, M. Horodecki, and P. Horodecki. Low-dimensional bound entanglement with one-way distillable cryptographic key. *IEEE Transactions on Information Theory*, 54(6):2621–2625, 2008.

[116] M. Horodecki, P. Horodecki, and R. Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996.

[117] M. Horodecki, P. Horodecki, and R. Horodecki. Mixed-state entanglement and distillation: is there a "bound" entanglement in nature? *Physical Review Letters*, 80(24):5239–5242, 1998.

[118] M. Horodecki, J. Oppenheim, and A. Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269(1):107–136, 2007.

[119] P. Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Physics Letters A*, 232(5):333–339, 1997.

[120] P. Horodecki. From entanglement witnesses to positive maps: towards optimal characterisation of separability. In A. Gonis and P. Turchi, editors, *Decoherence and its Implications in Quantum Computing and Information Transfer*, volume 182 of *NATO Science Series III: Computer and System Sciences*, pages 299–307. IOS Press, 2001.

[121] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of modern physics*, 81(865):865–942, 2009.

[122] R. Hudson and G. Moody. Locally normal symmetric states and an analogue of de Finetti's theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 33(4):343–351, 1976.

[123] L. Hughston, R. Jozsa, and W. Wootters. A complete classification of quantum ensembles having a given density matrix. *Physics Letters A*, 183(1):14–18, 1993.

[124] R. Jain. Distinguishing sets of quantum states. Unpublished manuscript. Available as arXiv.org e-Print quant-ph/0506205, 2005.

[125] A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.

[126] N. Johnston, D. Kribs, and V. Paulsen. Computing stabilized norms for quantum operations. *Quantum Information and Computation*, 9(1):16–35, 2009.

[127] R. Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315–2323, 1994.

[128] N. Killoran. *Entanglement quantification and quantum benchmarking of optical communication devices*. PhD thesis, University of Waterloo, 2012.

[129] A. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[130] A. Kitaev, A. Shen, and M. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002.

[131] O. Klein. Zur quantenmechanischen begründung des zweiten hauptsatzes der wärmelehre. *Zeitschrift für Physik*, 72(11–12):767–775, 1931.

[132] R. Klesse. A random coding based proof for the quantum coding theorem. *Open Systems & Information Dynamics*, 15(1):21–45, 2008.

[133] R. König and R. Renner. A de Finetti representation for finite symmetric quantum states. *Journal of Mathematical Physics*, 46(12):122108, 2005.

[134] K. Kraus. General state changes in quantum theory. *Annals of Physics*, 64:311–335, 1971.

[135] K. Kraus. *States, Effects, and Operations: Fundamental Notions of Quantum Theory*. Springer-Verlag, 1983.

[136] D. Kretschmann, D. Schlingemann, and R. Werner. The information-disturbance tradeoff and the continuity of Stinespring's representation. *IEEE Transactions on Information Theory*, 54(4):1708–1717, 2008.

[137] D. Kretschmann and R. Werner. Tema con variazioni: quantum channel capacity. *New Journal of Physics*, 6(1):26, 2004.

[138] D. Kribs. Quantum channels, wavelets, dilations and representations of $\mathcal{O}_n$. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 46:421–433, 2003.

[139] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[140] B. Kümmerer and H. Maassen. The essentially commutative dilations of dynamical semigroups on M_n. *Communications in Mathematical Physics*, 109(1):1–22, 1987.

[141] L. Landau. Das dämpfungsproblem in der wellenmechanik. *Zeitschrift für Physik*, 45:430–441, 1927.

[142] L. Landau and R. Streater. On Birkhoff's theorem for doubly stochastic completely positive maps of matrix algebras. *Linear Algebra and Its Applications*, 193:107–127, 1993.

[143] O. Lanford and D. Robinson. Mean entropy of states in quantum-statistical mechanics. *Journal of Mathematical Physics*, 9(7):1120–1125, 1968.

[144] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.

[145] P. Lévy. *Problémes concrets d'analyse fonctionelle*. Gauthier Villars, 1951.

[146] E. Lieb. Convex trace functions and the Wigner–Yanase–Dyson conjecture. *Advances in Mathematics*, 11(3):267–288, 1973.

[147] E. Lieb and M. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14(12):1938–1941, 1973.

[148] G. Lindblad. Expectation and entropy inequalities for finite quantum systems. *Communications in Mathematical Physics*, 39(2):111–119, 1974.

[149] G. Lindblad. A general no-cloning theorem. *Letters in Mathematical Physics*, 47(2):189–196, 1999.

[150] S. Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55(3):1613–1622, 1997.

[151] H.-K. Lo and S. Popescu. Concentrating entanglement by local actions:âĂČbeyond mean values. *Physical Review A*, 63(2):022301, 2001.

[152] M. Marcus. *Finite Dimensional Multilinear Algebra*, volume 1. Marcel Decker, 1973.

[153] M. Marcus. *Finite Dimensional Multilinear Algebra*, volume 2. Marcel Decker, 1975.

[154] A. Marshall, I. Olkin, and B. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer, second edition, 2011.

[155] B. Maurey and G. Pisier. Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. *Studia Mathematica*, 58(1):45–90, 1976.

[156] M. Mehta. *Random Matrices*. Elsevier, 2004.

[157] V. Mil'man. New proof of the theorem of A. Dvoretzky on intersections of convex bodies. *Functional Analysis and Its Applications*, 5(4):288–295, 1971.

[158] V. Milman and G. Schechtman. *Asymptotic theory of finite dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer, 1986.

[159] M. Naimark. On a representation of additive operator set functions. *Doklady Akademii Nauk SSSR*, 41:359–361, 1943.

[160] M. Nathanson. Distinguishing bipartitite orthogonal states using LOCC: Best and worst cases. *Journal of Mathematical Physics*, 46(6):062103, 2005.

[161] A. Nayak. *Lower bounds for Quantum Computation and Communication*. PhD thesis, University of California, Berkeley, 1999.

[162] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *40th Annual IEEE Symposium on Foundations of Computer Science*, pages 369–376, 1999.

[163] M. Nielsen. Conditions for a class of entanglement transformations. *Physical Review Letters*, 83(2):436–439, 1999.

[164] M. Nielsen. Probability distributions consistent with a mixed state. *Physical Review A*, 62(5):052308, 2000.

[165] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[166] M. Nielson. *Quantum Information Theory*. PhD thesis, University of New Mexico, 1998.

[167] K. Parthasarathy. Extremal decision rules in quantum hypothesis testing. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 2(4):557–568, 1999.

[168] V. Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2002.

[169] A. Peres. *Quantum Theory: Concepts and Methods*. Kluwer Academic Publishers, 1993.

[170] A. Peres. Separability criterion for density matrices. *Physical Review Letters*, 77(8):1413–1415, 1996.

[171] A. Peres and W. Wootters. Optimal detection of quantum information. *Physical Review Letters*, 66(9):1119–1122, 1991.

[172] D. Perez-Garcia, M. Wolf, D. Petz, and M. Ruskai. Contractivity of positive and trace-preserving maps under L_p norms. *Journal of Mathematical Physics*, 47(8):083506, 2006.

[173] M. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden–Day, 1964.

[174] E. Rains. Entanglement purification via separable superoperators. Unpublished manuscript. Available as arXiv.org e-Print quant-ph/9707002, 1997.

[175] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[176] R. Rosenkrantz, editor. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*. Kluwer Academic Publishers, 1989.

[177] B. Rosgen and J. Watrous. On the hardness of distinguishing mixed-state quantum computations. In *Proceedings of the 20th Annual Conference on Computational Complexity*, pages 344–354, 2005.

[178] W. Rudin. *Principles of Mathematical Analysis*. McGraw–Hill, 1964.

[179] B. Russo and H. Dye. A note on unitary operators in C*-algebras. *Duke Mathematical Journal*, 33(2):413–416, 1966.

[180] E. Schrödinger. Die gegenwärtige situation in der quantenmechanik. *Die Naturwissenschaften*, 23(48):807–812, 1935.

[181] E. Schrödinger. Die gegenwärtige situation in der quantenmechanik. *Die Naturwissenschaften*, 23(49):823–828, 1935.

[182] E. Schrödinger. Die gegenwärtige situation in der quantenmechanik. *Die Naturwissenschaften*, 23(50):844–849, 1935.

[183] E. Schrödinger. Discussion of probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(4):555–563, 1935.

[184] E. Schrödinger. Probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32(3):446–452, 1936.

[185] B. Schumacher. Quantum coding. *Physical Review A*, 51(4):2738–2747, 1995.

[186] B. Schumacher. Sending entanglement through noisy quantum channels. *Physical Review A*, 54(4):2614–2628, 1996.

[187] B. Schumacher and M. Nielsen. Quantum data processing and error correction. *Physical Review A*, 54(4):2629–2635, 1996.

[188] B. Schumacher and M. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, 1997.

[189] I. Schur. Über eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 22:9–20, 1923.

[190] J. Schur. Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 140:1–28, 1911.

[191] C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.

[192] P. Shor. Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics*, 246(3):453–472, 2004.

[193] B. Simon. *Trace ideals and their applications*, volume 35 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1979.

[194] G. Smith and J. Yard. Quantum communication with zero-capacity channels. *Science*, 321(5897):1812–1815, 2008.

[195] R. Smith. Completely bounded maps between C^*-algebras. *Journal of the London Mathematical Society*, 2(1):157–166, 1983.

[196] R. Spekkens and T. Rudolph. Degrees of concealment and bindingness in quantum bit commitment protocols. *Physical Review A*, 65(1):012310, 2001.

[197] W. F. Stinespring. Positive functions on C^*-algebras. *Proceedings of the American Mathematical Society*, 6(2):211–216, 1955.

[198] E. Størmer. Positive linear maps of operator algebras. *Acta Mathematica*, 110(1):233–278, 1963.

[199] M. Talagrand. *The Generic Chaining: Upper and Lower Bounds of Stochastic Processes*. Springer, 2006.

[200] B. Terhal and P. Horodecki. Schmidt number for density matrices. *Physical Review A*, 61(4):040301, 2000.

[201] R. Timoney. Computing the norms of elementary operators. *Illinois Journal of Mathematics*, 47(4):1207–1226, 2003.

[202] S. Tregub. Bistochastic operators on finite-dimensional von Neumann algebras. *Izvestiya Vysshikh Uchebnykh Zavedenii Matematika*, 30(3):75–77, 1986.

[203] M. Tribus and E. McIrvine. Energy and information. *Scientific American*, 225(3):179–188, 1971.

[204] J. Trimmer. The present situation in quantum mechanics: a translation of Schrödinger's "cat paradox" paper. *Proceedings of the American Philosophical Society*, 124(5):323–338, 1980.

[205] B. Tsirel'son. Quantum analogues of the Bell inequalities. the case of two spatially separated domains. *Journal of Soviet Mathematics*, 36:557–570, 1987.

[206] A. Uhlmann. Sätze über dichtematrizen. *Wissenschaftliche Zeitschrift der Karl-Marx-Universitat Leipzig. Mathematisch-naturwissenschaftliche Reihe*, 20(4/5):633–653, 1971.

[207] A. Uhlmann. Endlich-dimensionale dichtematrizen I. *Wissenschaftliche Zeitschrift der Karl-Marx-Universitat Leipzig. Mathematisch-naturwissenschaftliche Reihe*, 21(4):421–452, 1972.

[208] A. Uhlmann. Endlich-dimensionale dichtematrizen II. *Wissenschaftliche Zeitschrift der Karl-Marx-Universitat Leipzig. Mathematisch-naturwissenschaftliche Reihe*, 22(2):139–177, 1973.

[209] A. Uhlmann. The "transition probability" in the state space of a $*$-algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976.

[210] A. Uhlmann. Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Communications in Mathematical Physics*, 54(1):21–32, 1977.

[211] H. Umegaki. Conditional expectations in an operator algebra IV (entropy and information). *Kodai Mathematical Seminar Reports*, 14(2):59–85, 1962.

[212] V. Vedral, M. Plenio, M. Rippin, and P. Knight. Quantifying entanglement. *Physical Review Letters*, 78(12):2275–2278, 1997.

[213] J. von Neumann. Thermodynamik quantenmechanischer gesamtheiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen*, 1(11):273–291, 1927.

[214] J. von Neumann. Wahrscheinlichkeitstheoretischer aufbau der mechanik. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen*, 1(11):245–272, 1927.

[215] J. von Neumann. Zur algebra der funktionaloperationen und theorie der normalen operatoren. *Mathematische Annalen*, 102(1):370–427, 1930.

[216] J. von Neumann. Die einfuhrung analytischer parameter in topologischen gruppen. *Annals of Mathematics (Second Series)*, 34(1):170–179, 1933.

[217] J. von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1955. Originally published in German in 1932 as *Mathematische Grundlagen der Quantenmechanik*.

[218] J. Walgate, A. Short, L. Hardy, and V. Vedral. Local distinguishability of multipartite orthogonal quantum states. *Physical Review Letters*, 85(23):4972–4975, 2000.

[219] J. Watrous. Notes on super-operator norms induced by Schatten norms. *Quantum Information and Computation*, 5(1):58–68, 2005.

[220] J. Watrous. Distinguishing quantum operations having few Kraus operators. *Quantum Information and Computation*, 8(9):819–833, 2008.

[221] J. Watrous. Mixing doubly stochastic quantum channels with the completely depolarizing channel. *Quantum Information and Computation*, 9(5/6):406–413, 2009.

[222] J. Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11), 2009.

[223] J. Watrous. Simpler semidefinite programs for completely bounded norms. *Chicago Journal of Theoretical Computer Science*, 2013:8, 2013.

[224] A. Weil. *L'intégration dans les groupes topologiques et ses applications*. Hermann, second edition, 1979. Originally published in 1940.

[225] R. Werner. Quantum states with Einstein–Podolsky–Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40(8):4277–4281, 1989.

[226] R. Werner. Optimal cloning of pure states. *Physical Review A*, 58(3):1827–1832, 1998.

[227] R. Werner. All teleportation and dense coding schemes. *Journal of Physics A: Mathematical and General*, 34(35):7081–7094, 2001.

[228] H. Weyl. *The Theory of Groups and Quantum Mechanics*. Dover Publications, 1950. Originally published in German in 1929.

[229] M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.

[230] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999.

[231] H. Wolkowicz, R. Saigal, and L. Vandenberge, editors. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, 2000.

[232] S. Woronowicz. Positive maps of low dimensional matrix algebras. *Reports on Mathematical Physics*, 10(2):165–183, 1976.

[233] D. Yang, M. Horodecki, R. Horodecki, and B. Synak-Radtke. Irreversibility for all bound entangled states. *Physical Review Letters*, 95(19):190501, 2005.

[234] H. Yuen, R. Kennedy, and M. Lax. On optimal quantum receivers for digital signal detection. *Proceedings of the IEEE*, 58(10):1770–1773, 1970.

[235] H. Yuen, R. Kennedy, and M. Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Transactions on Information Theory*, 21(2):125–134, 1975.

[236] V. Zarikian. Alternating-projection algorithms for operator-theoretic calculation. *Linear Algebra and Its Applications*, 419(2–3):710–734, 2006.

[237] K. Życzkowski, P. Horodecki, A. Sanpera, and M. Lewenstein. Volume of the set of separable states. *Physical Review A*, 58(2):883–892, 1998.