Chapter 2
Modeling Quantum Information

Abstract Classical as well as quantum information is stored in physical systems, or "information is inevitably physical" as Rolf Landauer famously said. These physical systems are ultimately governed by the laws of quantum mechanics. In this chapter we quickly review the relevant mathematical foundations of quantum theory and introduce notational conventions that will be used throughout the book.

In particular we will discuss concepts of functional and matrix analysis as well as linear algebra that will be of use later. We consider general separable Hilbert spaces in this chapter, even though in the rest of the book we restrict our attention to the finite-dimensional case. This digression is useful because it motivates the notation we use throughout the book, and it allows us to distinguish between the mathematical structure afforded by quantum theory and the additional structure that is only present in the finite-dimensional case.

Our notation is summarized in Sect. 2.1 and the remainder of this chapter can safely be skipped by expert readers. The presentation here is compressed and we omit proofs. We instead refer to standard textbooks (see Sect. 2.7 for some references) for a more comprehensive treatment.

2.1 General Remarks on Notation

The notational conventions for this book are summarized in Table 2.1. The table includes references to the sections where the corresponding concepts are introduced. Throughout this book we are careful to distinguish between linear operators (e.g. events and Kraus operators) and functionals on the linear operators (e.g. states), which are also represented as linear operators (e.g. density operators). This distinction is inspired by the study of infinite-dimensional systems where these objects do not necessarily have the same mathematical structure, but it is also helpful in the finite-dimensional setting.[1]

[1] For example, it sheds light on the fact that we use the operator norm for ordinary linear operators and its dual norm, the trace norm, for density operators.

M. Tomamichel, *Quantum Information Processing with Finite Resources*,
SpringerBriefs in Mathematical Physics, DOI 10.1007/978-3-319-21891-5_2

Table 2.1 Overview of notational conventions

Symbol	Variants	Description	Section
$\mathbb{R}, \mathbb{C}$	$\mathbb{R}_+$	Real and complex fields (and non-negative reals)	
$\mathbb{N}$		Natural numbers	
log, exp	ln, e	Logarithm (to unspecified basis), and its inverse, the exponential function (natural logarithm and Euler's constant)	
$\mathscr{H}$	$\mathscr{H}_{AB}, \mathscr{H}_X$	Hilbert spaces (for joint system AB and system X)	2.2.1
$\langle\cdot\|, \|\cdot\rangle$		Bra and ket	
$\mathrm{Tr}(\cdot)$	Tr_A	Trace (partial trace)	2.3.1
$\otimes$	$(\cdot)^{\otimes n}$	Tensor product (n-fold tensor product)	2.4.1
$\oplus$		Direct sum for block diagonal operators	2.2.2
$A \ll B$		A is dominated by B, i.e. kernel of A contains kernel of B	
$A \perp B$		A and B are orthogonal, i.e. $AB = BA = 0$	
$\mathscr{L}$	$\mathscr{L}(A, B)$	Bounded linear operators (from $\mathscr{H}_A$ to $\mathscr{H}_B$)	2.2.1
$\mathscr{L}^\dagger$	$\mathscr{L}^\dagger(B)$	Self-adjoint operators (acting on $\mathscr{H}_B$)	
$\mathscr{P}$	$\mathscr{P}(CD)$	Positive semi-definite operators (acting on $\mathscr{H}_{CD}$)	
$\{A \geq B\}$		Projector on subspace where $A - B$ is non-negative	
$\\|\cdot\\|$		Operator norm	2.2.1
$\mathscr{L}_\bullet$	$\mathscr{L}_\bullet(E)$	Contractions in $\mathscr{L}$ (acting on $\mathscr{H}_E$)	
$\mathscr{P}_\bullet$	$\mathscr{P}_\bullet(A)$	Contractions in $\mathscr{P}$ (corresponding to events on A)	2.2.2
I	I_Y	Identity operator (acting on $\mathscr{H}_Y$)	
$\langle\cdot,\cdot\rangle$		Hilbert-Schmidt inner product	2.3.1
$\mathscr{T}$	$\mathscr{T} \equiv \mathscr{L}^{\ddagger}$	Trace-class operators representing linear functionals	
$\mathscr{S}$	$\mathscr{S} \equiv \mathscr{P}^{\ddagger}$	Operators representing positive functionals	
$\\|\cdot\\|_*$	$\mathrm{Tr}\|\cdot\|$	Trace norm on functionals	2.3.1
$\mathscr{S}_\bullet$	$\mathscr{S}_\bullet(A)$	Sub-normalized density operators (on A)	2.3.2
$\mathscr{S}_\circ$	$\mathscr{S}_\circ(B)$	Normalized density operators, or states (on B)	
π	π_A	Fully mixed state (on A), in finite dimensions	2.3.2
ψ	ψ_{AB}	Maximally entangled state (between A and B), in finite dimensions	2.4.2
CB	$\mathrm{CB}(A, B)$	Completely bounded maps (from $\mathscr{L}(A)$ to $\mathscr{L}(B)$)	2.6.1
CP		Completely positive maps	2.6.2
CPTP	CPTNI	Completely positive trace-preserving (trace-non-increasing) map	
$\\|\cdot\\|_+$	$\\|\cdot\\|_p$	Positive cone dual norm (Schatten p-norm)	3.1
$\Delta(\cdot,\cdot)$		Generalized trace distance for sub-normalized states	3.2
$F(\cdot,\cdot)$	$F_*(\cdot,\cdot)$	Fidelity (generalized fidelity for sub-normalized states)	3.3
$P(\cdot,\cdot)$		Purified distance for sub-normalized states	3.4

‡This equivalence only holds if the underlying Hilbert space is finite-dimensional

We do not specify a particular basis for the logarithm throughout this book, and simply use exp to denote the inverse of log.[2] The natural logarithm is denoted by ln.

We label different physical systems by capital Latin letters A, B, C, D, and E, as well as X, Y, and Z which are specifically reserved for classical systems. The label thus always determines if a system is quantum or classical. We often use these labels as subscripts to guide the reader by indicating which system a mathematical object belongs to. We drop the subscripts when they are evident in the context of an expression (or if we are not talking about a specific system). We also use the capital Latin letters L, K, H, M, and N to denote linear operators, where the last two are reserved for positive semi-definite operators. The identity operator is denoted I. Density operators, on the other hand, are denoted by lowercase Greek letters ρ, τ, σ, and ω. We reserve π and ψ for the fully mixed state and the maximally entangled state, respectively. Calligraphic letters are used to denote quantum channels and other maps acting on operators.

2.2 Linear Operators and Events

For our purposes, a *physical system* is fully characterized by the set of events that can be observed on it. For classical systems, these events are traditionally modeled as a σ-algebra of subsets of the sample space, usually the power set in the discrete case. For quantum systems the structure of events is necessarily more complex, even in the discrete case. This is due to the non-commutative nature of quantum theory: the union and intersection of events are generally ill-defined since it matters in which order events are observed.

Let us first review the mathematical model used to describe events in quantum mechanics (as positive semi-definite operators on a Hilbert space). Once this is done, we discuss physical systems carrying quantum and classical information.

2.2.1 Hilbert Spaces and Linear Operators

For concreteness and to introduce the notation, we consider two physical systems A and B as examples in the following. We associate to A a separable *Hilbert space* $\mathscr{H}_A$ over the field $\mathbb{C}$, equipped with an *inner product* $\langle \cdot, \cdot \rangle : \mathscr{H}_A \times \mathscr{H}_A \to \mathbb{C}$. In the finite-dimensional case, this is simply a complex inner product space, but we will follow a tradition in quantum information theory and call $\mathscr{H}_A$ a Hilbert space also in this case. Analogously, we associate the Hilbert space $\mathscr{H}_B$ to the physical system B.

[2] The reader is invited to think of $\log(x)$ as the binary logarithm of x and, consequently, $\exp(x) = 2^x$, as is customary in quantum information theory.

Linear Operators

Our main object of study are *linear operators* acting on the system's Hilbert space. We consistently use upper-case Latin letters to denote such linear operators. More precisely, we consider the set of *bounded linear operators* from $\mathscr{H}_A$ to $\mathscr{H}_B$, which we denote by $\mathscr{L}(A, B)$. Bounded here refers to the *operator norm* induced by the Hilbert space's inner product.

The **operator norm** on $\mathscr{L}(A, B)$ is defined as

$$\|\cdot\| : \quad L \mapsto \sup\left\{\sqrt{\langle Lv, Lv\rangle_B} \ : \ v \in \mathscr{H}_A,\ \langle v, v\rangle_A \leq 1\right\}. \tag{2.1}$$

For all $L \in \mathscr{L}(A, B)$, we have $\|L\| < \infty$ by definition. A linear operator is continuous if and only if it is bounded.[3] Let us now summarize some important concepts and notation that we will frequently use throughout this book.

- The *identity* operator on $\mathscr{H}_A$ is denoted I_A.
- The *adjoint* of a linear operator $L \in \mathscr{L}(A, B)$ is the unique operator $L^\dagger \in \mathscr{L}(B, A)$ that satisfies $\langle w, Lv\rangle_B = \langle L^\dagger w, v\rangle_A$ for all $v \in \mathscr{H}_A$, $w \in \mathscr{H}_B$. Clearly, $(L^\dagger)^\dagger = L$.
- For scalars $\alpha \in \mathbb{C}$, the adjoint corresponds to the complex conjugate, $\alpha^\dagger = \overline{\alpha}$.
- We find $(LK)^\dagger = K^\dagger L^\dagger$ by applying the definition twice.
- The *kernel* of a linear operator $L \in \mathscr{L}(A, B)$ is the subspace of $\mathscr{H}_A$ spanned by vectors $v \in \mathscr{H}_A$ satisfying $Lv = 0$. The *support* of L is its orthogonal complement in $\mathscr{H}_A$ and the *rank* is the cardinality of the support. Finally, the image of L is the subspace of $\mathscr{H}_B$ spanned by vectors $w \in \mathscr{H}_B$ such that $w = Lv$ for some $v \in \mathscr{H}_A$.
- For operators $K, L \in \mathscr{L}(A)$ we say that L is *dominated* by K if the kernel of K is contained in the kernel of L. Namely, we write $L \ll K$ if and only if

$$K\,|v\rangle_A = 0 \implies L\,|v\rangle_A = 0 \qquad \text{for all} \quad v \in \mathscr{H}_A. \tag{2.3}$$

- We say $K, L \in \mathscr{L}(A)$ are *orthogonal* (denoted $K \perp L$) if $KL = LK = 0$.
- We call a linear operator $U \in \mathscr{L}(A, B)$ an *isometry* if it preserves the inner product, namely if $\langle Uv, Uw\rangle_B = \langle v, w\rangle_A$ for all $v, w \in \mathscr{H}_A$. This holds if $U^\dagger U = I_A$.

---

[3] *Relation to Operator Algebras:* Let us note that $\mathscr{L}(A, B)$ with the norm $\|\cdot\|$ is a Banach space over $\mathbb{C}$. Furthermore, the operator norm satisfies

$$\|L\|^2 = \|L^\dagger\|^2 = \|L^\dagger L\| \quad \text{and} \quad \|LK\| \leq \|L\| \cdot \|K\|. \tag{2.2}$$

for any $L \in \mathscr{L}(A, B)$ and $K \in \mathscr{L}(B, A)$. The inequality states that the norm is *sub-multiplicative*.

The above properties of the norm imply that the space $\mathscr{L}(A)$ is (weakly) closed under multiplication and the adjoint operation. In fact, $\mathscr{L}(A)$ constitutes a (Type I factor) von Neumann algebra or C^* algebra. Alternatively, we could have started our considerations right here by postulating a Type 1 von Neumann algebra as the fundamental object describing individual physical systems, and then deriving the Hilbert space structure as a consequence.

- An isometry is an example of a *contraction*, i.e. an operator $L \in \mathscr{L}(A, B)$ satisfying $\|L\| \leq 1$. The set of all such contractions is denoted $\mathscr{L}_\bullet(A, B)$. Here the bullet '•' in the subscript of $\mathscr{L}_\bullet(A, B)$ simply illustrates that we restrict $\mathscr{L}(A, B)$ to the unit ball for the norm $\|\cdot\|$.

For any $L \in \mathscr{L}(A)$, we denote by L^{-1} its Moore-Penrose *generalized inverse* or pseudoinverse [130] (which always exists in finite dimensions). In particular, the generalized inverse satisfies $LL^{-1}L = L$ and $L^{-1}LL^{-1} = L^{-1}$. If $L = L^\dagger$, the generalized inverse is just the usual inverse evaluated on the operator's support.

Bras, Kets and Orthonormal Bases

We use the *bra-ket notation* throughout this book. For any vector $v_A \in \mathscr{H}_A$, we use its *ket*, denoted $|v\rangle_A$, to describe the embedding

$$|v\rangle_A : \mathbb{C} \to \mathscr{H}_A, \quad \alpha \mapsto \alpha v_A. \tag{2.4}$$

Similarly, we use its *bra*, denoted $\langle v|_A$, to describe the functional

$$\langle v|_A : \mathscr{H}_A \to \mathbb{C}, \quad w_A \mapsto \langle v, w\rangle_A. \tag{2.5}$$

It is natural to view kets as linear operators from $\mathbb{C}$ to $\mathscr{H}_A$ and bras as linear operators from $\mathscr{H}_A$ to $\mathbb{C}$. The above definitions then imply that

$$|Lv\rangle_A = L\,|v\rangle_A\,, \quad \langle Lv|_A = \langle v|_A\, L^\dagger, \quad \text{and} \quad \langle v|_A = |v\rangle_A^\dagger\,. \tag{2.6}$$

Moreover, the inner product can equivalently be written as $\langle w, Lv\rangle_B = \langle w|_B\, L|v\rangle_A$. Conjugate symmetry of the inner product then corresponds to the relation

$$\overline{\langle w|_B L|v\rangle_A} = \langle v|_A L^\dagger |w\rangle_B. \tag{2.7}$$

As a further example, we note that $|v\rangle_A$ is an isometry if and only if $\langle v|v\rangle_A = 1$.

In the following we will work exclusively with linear operators (including bras and kets) and we will not use the underlying vectors (the elements of the Hilbert space) or the inner product of the Hilbert space anymore.

We now restrict our attention to the space $\mathscr{L}(A) := \mathscr{L}(A, A)$ of bounded linear operators acting on $\mathscr{H}_A$. An operator $U \in \mathscr{L}(A)$ is *unitary* if U and $U^\dagger$ are isometries. An *orthonormal basis* (ONB) of the system A (or the Hilbert space $\mathscr{H}_A$) is a set of vectors $\{e_x\}_x$, with $e_x \in \mathscr{H}_A$, such that

$$\langle e_x | e_y \rangle_A = \delta_{x,y} := \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases} \quad \text{and} \quad \sum_x |e_x\rangle\langle e_x|_A = I_A. \tag{2.8}$$

We denote the dimension of $\mathscr{H}_A$ by d_A if it is finite and note that the index x ranges over d_A distinct values. For general separable Hilbert spaces x ranges over any countable set. (We do not usually specify such index sets explicitly.) Various ONBs exist and are related by unitary operators: if $\{e_x\}_x$ is an ONB then $\{Ue_x\}_x$ is too, and, furthermore, given two ONBs there always exists a unitary operator mapping one basis to the other, and vice versa.

Positive Semi-Definite Operators

A special role is played by operators that are self-adjoint and positive semi-definite. We call an operator $H \in \mathscr{L}(A)$ *self-adjoint* if it satisfies $H = H^\dagger$, and the set of all self-adjoint operators in $\mathscr{L}(A)$ is denoted $\mathscr{L}^\dagger(A)$. Such self-adjoint operators have a spectral decomposition,

$$H = \sum_x \lambda_x \, |e_x\rangle\langle e_x| \tag{2.9}$$

where $\{\lambda_x\}_x \subset \mathbb{R}$ are called *eigenvalues* and $\{|e_x\rangle\}_x$ is an orthonormal basis with *eigenvectors* $|e_x\rangle$. The set $\{\lambda_x\}_x$ is also called the *spectrum* of H, and it is unique.

Finally we introduce the set $\mathscr{P}(A)$ of *positive semi-definite* operators in $\mathscr{L}(A)$. An operator $M \in \mathscr{L}(A)$ is positive semi-definite if and only if $M = L^\dagger L$ for some $L \in \mathscr{L}(A)$, so in particular such operators are self-adjoint and have non-negative eigenvalues. Let us summarize some important concepts and notation concerning self-adjoint and positive semi-definite operators here.

- We call $P \in \mathscr{P}(A)$ a *projector* if it satisfies $P^2 = P$, i.e. if it has only eigenvalues 0 and 1. The identity I_A is a projector.
- For any $K, L \in \mathscr{L}^\dagger(A)$, we write $K \geq L$ if $K - L \in \mathscr{P}(A)$. Thus, the relation '$\geq$' constitutes a partial order on $\mathscr{L}(A)$.
- For any $G, H \in \mathscr{L}^\dagger(A)$, we use $\{G \geq H\}$ to denote the projector onto the subspace corresponding to non-negative eigenvalues of $G - H$. Analogously, $\{G < H\} = I - \{G \geq H\}$ denotes the projector onto the subspace corresponding to negative eigenvalues of $G - H$.

Matrix Representation and Transpose

Linear operators in $\mathscr{L}(A, B)$ can be conveniently represented as matrices in $\mathbb{C}^{d_A} \times \mathbb{C}^{d_B}$. Namely for any $L \in \mathscr{L}(A, B)$, we can write

$$L = \sum_{x,y} |f_y\rangle\langle f_y|_B \, L|e_x\rangle\langle e_x|_A = \sum_{x,y} \langle f_y|L|e_x\rangle \cdot |f_y\rangle\langle e_x|, \tag{2.10}$$

where $\{e_x\}_x$ is an ONB of A and $\{f_y\}_y$ an ONB of B. This decomposes L into elementary operators $|f_y\rangle\langle e_x| \in \mathscr{L}_\bullet(A, B)$ and the matrix with entries $[L]_{yx} = \langle f_y|L|e_x\rangle$.

Moreover, there always exists a choice of the two bases such that the resulting matrix is diagonal. For such a choice of bases, we find the *singular value decomposition* $L = \sum_x s_x |f_x\rangle\langle e_x|$, where $\{s_x\}_x$ with $s_x \geq 0$ are called the singular values of L. In particular, for self-adjoint operators, we can choose $|f_x\rangle = |e_x\rangle$ and recover the eigenvalue decomposition with $s_x = |\lambda_x|$.

The *transpose* of L with regards to the bases $\{e_x\}$ and $\{f_y\}$ is defined as

$$L^T := \sum_{x,y} \langle f_y|L|e_x\rangle \cdot |e_x\rangle\langle f_y|, \quad L^T \in \mathscr{L}(B, A). \tag{2.11}$$

Importantly, in contrast to the adjoint, the transpose is only defined with regards to a particular basis. Also contrast (2.11) with the matrix representation of $L^\dagger$,

$$L^\dagger = \sum_{x,y} \big(\langle f_y|L|e_x\rangle\big)^\dagger \cdot |e_x\rangle\langle f_y| = \sum_{x,y} \langle e_x|L^\dagger|f_y\rangle \cdot |e_x\rangle\langle f_y| = \overline{L}^T. \tag{2.12}$$

Here, $\overline{L}$ denotes the complex conjugate, which is also basis dependent.

2.2.2 Events and Measures

We are now ready to attach physical meaning to the concepts introduced in the previous section, and apply them to physical systems carrying quantum information.

Observable **events** on a quantum system A correspond to operators in the unit ball of $\mathscr{P}(A)$, namely the set

$$\mathscr{P}_\bullet(A) := \{M \in \mathscr{L}(A) : \ 0 \leq M \leq I\}. \tag{2.13}$$

(The bullet '•' indicates that we restrict to the unit ball of the norm $\|\cdot\|$.)

Two events $M, N \in \mathscr{P}_\bullet(A)$ are called *exclusive* if $M + N$ is an event in $\mathscr{P}_\bullet(A)$ as well. In this case, we call $M + N$ the *union* of the events M and N. A complete set of mutually exclusive events that sum up to the identity is called a *positive operator valued measure* (POVM). More generally, for any measurable space $(\mathscr{X}, \Sigma)$ with Σ a σ-algebra, a POVM is a function

$$O_A : \Sigma \to \mathscr{P}_\bullet(A) \quad \text{with} \quad O_A(\mathscr{X}) = I_A \tag{2.14}$$

that is σ-additive, meaning that $O_A(\bigcup_i \mathscr{X}_i) = \sum_i O_A(\mathscr{X}_i)$ for mutually disjoint subsets $\mathscr{X}_i \subset \mathscr{X}$. This definition is too general for our purposes here, and we will restrict our attention to the case where $\mathscr{X}$ is discrete and Σ the power set of $\mathscr{X}$. In that case the POVM is fully determined if we associate mutually exclusive events to each $x \in \mathscr{X}$.

A function $x \mapsto M_A(x)$ with $M_A(x) \in \mathscr{P}_\bullet(A)$, $\sum_x M_A(x) = I_A$ is called a **positive operator valued measure** (POVM) on A.

We assume that x ranges over a countable set for this definition, and we will in fact not discuss measurements with continuous outcomes in this book. We call $x \mapsto M_A(x)$ a *projective* measure if all $M_A(x)$ are projectors, and we call it *rank-one* if all $M_A(x)$ have rank one.

Structure of Classical Systems

Classical systems have the distinguishing property that all events commute.

To model a classical system X in our quantum framework, we restrict $\mathscr{P}_\bullet(X)$ to a set of events that commute. These are diagonalized by a common ONB, which we call the *classical basis* of X. For simplicity, the classical basis is denoted $\{x\}_x$ and the corresponding kets are $|x\rangle_X$. (To avoid confusion, we will call the index y or z instead of x if the systems Y and Z are considered instead.)

Every $M \in \mathscr{P}_\bullet(X)$ on a classical system can be written as

$$M = \sum_x M(x)\,|x\rangle\langle x|_X = \bigoplus_x M(x), \quad \text{where} \quad 0 \le M(x) \le 1. \tag{2.15}$$

Instead of writing down the basis projectors, $|x\rangle\langle x|$, we sometimes employ the direct sum notation to illustrate the block-diagonal structure of such operators. In the following, whenever we introduce a classical event M on X we also implicitly introduce the function $M(x)$, and vice versa.

This definition of "classical" events still goes beyond the usual classical formalism of discrete probability theory. In the usual formalism, M represents a subset of the sample space (an element of its σ-algebra), and thus corresponds to a projector in our language, with $M(x) \in \{0, 1\}$ indicating if x is in the set. Our formalism, in contrast, allows to model probabilistic events, i.e. the event M occurs at most with probability $M(x) \in [0, 1]$ even if the state is deterministically x.[4]

2.3 Functionals and States

States of a physical system are functionals on the set of bounded linear operators that map events to the probability that the respective event occurs.

[4]This generalization is quite useful as it, for example, allows us to see the optimal (probabilistic) Neyman-Pearson test as an event.

Continuous linear functionals can be represented as trace-class operators. This then allows us to introduce states for quantum and classical systems.

2.3.1 Trace and Trace-Class Operators

The most fundamental linear functional is the *trace*. For any orthonormal basis $\{e_x\}_x$ of A, we define the trace over A as

$$\mathrm{Tr}_A(\cdot) : \mathscr{L}(A) \to \mathbb{C}, \quad L \mapsto \sum_x \langle e_x | L | e_x \rangle_A . \tag{2.16}$$

Note that $\mathrm{Tr}(L)$ is finite if $d_A < \infty$ or more generally if L is *trace-class*, as we will see below. The trace is cyclic, namely we have

$$\mathrm{Tr}_A(KL) = \mathrm{Tr}_B(LK) \tag{2.17}$$

for any two operators $L \in \mathscr{L}(A, B)$, $K \in \mathscr{L}(B, A)$ when KL and LK are trace-class. Thus, in particular, for any $L \in \mathscr{L}(A)$, we have $\mathrm{Tr}_A(L) = \mathrm{Tr}_B(ULU^\dagger)$ for any *isometry* $U \in \mathscr{L}(A, B)$, which shows that the particular choice of basis used for the definition of the trace in (2.16) is irrelevant. Finally, we have $\mathrm{Tr}(L^\dagger) = \overline{\mathrm{Tr}(L)}$.

Trace-Class Operators

Using the trace, continuous linear functionals can be conveniently represented as elements of the dual Banach space of $\mathscr{L}(A)$, namely the space of linear operators on $\mathscr{H}_A$ with bounded *trace norm*.

The **trace norm** on $\mathscr{L}(A)$ is defined as

$$\|\cdot\|_* : \quad \xi \mapsto \mathrm{Tr}\,|\xi| = \mathrm{Tr}\left(\sqrt{\xi^\dagger \xi}\right). \tag{2.18}$$

Operators $\xi \in \mathscr{L}(A)$ with $\|\xi\|_* < \infty$ are called **trace-class operators**.

We denote the subspace of $\mathscr{L}(A)$ consisting of trace-class operators by $\mathscr{T}(A)$ and we use lower-case Greek letters to denote elements of $\mathscr{T}(A)$. In infinite dimensions $\mathscr{T}(A)$ is a proper subspace of $\mathscr{L}(A)$. In finite dimensions $\mathscr{L}(A)$ and $\mathscr{T}(A)$ coincide, but we will use this convention to distinguish between linear operators and linear operators representing functionals nonetheless.

For every trace-class operator $\xi \in \mathscr{T}(A)$, we define the functional $F_\xi(L) := \langle \xi, L \rangle$ using the sesquilinear form

$$\langle \cdot, \cdot \rangle : \mathscr{T}(A) \times \mathscr{L}(A) \to \mathbb{C}, \quad (\xi, L) \mapsto \mathrm{Tr}(\xi^\dagger L). \tag{2.19}$$

This form is continuous in both $\mathscr{L}(A)$ and $\mathscr{T}(A)$ with regards to the respective norms on these spaces, which is a direct consequence of Hölder's inequality $|\mathrm{Tr}(\xi^\dagger L)| \leq \|\xi\|_* \cdot \|L\|$.[5] In finite dimensions it is also tempting to view $\mathscr{L}(A) = \mathscr{T}(A)$ as a Hilbert space with $\langle \cdot, \cdot \rangle$ as its inner product, the *Hilbert-Schmidt inner product*. Finally, *positive functionals* map $\mathscr{P}(A)$ onto the positive reals. Since $\mathrm{Tr}(\omega M) \geq 0$ for all $M \geq 0$ if and only if $\omega \geq 0$, we find that positive functionals correspond to positive semi-definite operators in $\mathscr{T}(A)$, and we denote these by $\mathscr{S}(A)$.

2.3.2 States and Density Operators

A *state* of a physical system A is a functional that maps events $M \in \mathscr{P}_\bullet(A)$ to the respective probability that M is observed. We want the probability of the union of two mutually exclusive events to be additive, and thus such functionals must be linear. Furthermore, we require them to be continuous with regards to small perturbations of the events. Finally, they ought to map events into the interval [0, 1], hence they must also be positive and normalized.

Based on the discussion in the previous section, we can conveniently parametrize all functionals corresponding to states as follows. We define the set of *sub-normalized density operators* as trace-class operators in the unit ball,

$$\mathscr{S}_\bullet(A) := \{\rho_A \in \mathscr{T}(A) : \ \rho_A \geq 0 \ \wedge \ \mathrm{Tr}(\rho_A) \leq 1\}. \tag{2.21}$$

Here the bullet '•' refers to the unit ball in the norm $\| \cdot \|_*$. (This norm simply corresponds to the trace for positive semi-definite operators.)

For any operator $\rho_A \in \mathscr{S}_\bullet(A)$, we define the functional

$$\Pr_\rho(\cdot) : \ \mathscr{P}_\bullet(A) \to [0, 1], \quad M \mapsto \langle \rho_A, M \rangle = \mathrm{Tr}(\rho_A M), \tag{2.22}$$

which maps events to the probability that the event occurs.

[5] Note also that the norms $\| \cdot \|$ and $\| \cdot \|_*$ are dual with regards to this form, namely we have

$$\|\xi\|_* = \sup\left\{ |\langle \xi, L \rangle| : \ L \in \mathscr{L}_\bullet(A) \right\}. \tag{2.20}$$

The trace norm is thus sometimes also called the *dual norm*.

This is an expression of Born's rule, and often taken as an axiom of quantum mechanics. Here it is just a natural way to map events to probabilities. We call such operators ρ_A density operators.

It is often prudent to further require that the union of all events in a POVM, namely the event I, has probability 1. This leads us to normalized density operators:

Quantum states are represented as **normalized density operators** in

$$\mathscr{S}_\circ(A) := \{\rho_A \in \mathscr{T}(A) : \ \rho_A \geq 0 \ \wedge \ \mathrm{Tr}(\rho_A) = 1\}, \tag{2.23}$$

(The circle '∘' indicates that we restrict to the unit sphere of the norm $\|\cdot\|_*$.)

In the following we will use the expressions state and density operator interchangeably. We also use the set $\mathscr{S}$ which contains all positive semi-definite operators, if there is no need for normalization.

States form a convex set, and a state is called *mixed* if it lies in the interior of this set. The fully mixed state (in finite dimensions) is denoted $\pi_A := I_A/d_A$. On the other hand, states on the boundary are called *pure*. Pure states are represented by density operators with rank one, and can be written as $\phi_A = |\phi\rangle\langle\phi|_A$ for some $\phi \in \mathscr{H}_A$. With a slight abuse of nomenclature, we often call the corresponding ket, $|\phi\rangle_A$, a state.

Probability Mass Functions

The structure of density operators simplifies considerably for classical systems. We are interested in evaluating the probabilities for events of the form (2.15). Hence, for any $\rho_X \in \mathscr{S}_\circ(X)$, we find

$$\Pr_\rho(M) = \mathrm{Tr}(\rho_X M) = \sum_x M(x) \langle x| \rho_X |x\rangle_X = \sum_x M(x)\rho(x), \tag{2.24}$$

where we defined $\rho_X(x) = \langle x| \rho_X |x\rangle_X$. We thus see that it suffices to consider states of the following form:

States $\rho_X \in \mathscr{S}_\circ(X)$ on a classical system X have the form

$$\rho_X = \sum_x \rho(x) \, |x\rangle\langle x|_X \,, \quad \text{where} \quad \rho(x) \geq 0, \quad \sum_x \rho(x) = 1. \tag{2.25}$$

where $\rho(x)$ is called a **probability mass function**.

Moreover, if $\rho_X \in \mathscr{S}_\bullet(X)$ is a sub-normalized density operator, we require that $\sum_x \rho(x) \leq 1$ instead of the equality. Again, whenever we introduce a density operator ρ_X on X, we implicitly also introduce the function $\rho(x)$, and vice versa.

2.4 Multi-partite Systems

A joint system AB is modeled using bounded linear operators on a tensor product of Hilbert spaces, $\mathscr{H}_{AB} := \mathscr{H}_A \otimes \mathscr{H}_B$. The respective set of bounded linear operators is denoted $\mathscr{L}(AB)$ and the events on the joint systems are thus the elements of $\mathscr{P}_\bullet(AB)$. Analogously, all the other sets of operators defined in the previous sections are defined analogously for the joint system.

2.4.1 Tensor Product Spaces

For every $v \in \mathscr{H}_{AB}$ on the joint system AB, there exist two ONBs, $\{e_x\}_x$ on A and $\{f_y\}_y$ on B, as well as a unique set of positive reals, $\{\lambda_x\}_x$, such that we can write

$$|v\rangle_{AB} = \sum_x \sqrt{\lambda_x}\; |e_x\rangle_A \otimes |f_x\rangle_B. \tag{2.26}$$

This is called the *Schmidt decomposition* of v. The convention to use a square root is motivated by the fact that the sequence $\{\sqrt{\lambda_x}\}_x$ is square summable, i.e. $\sum_x \lambda_x < \infty$. Note also that $\{e_x \otimes f_y\}_{x,y}$ can be extended to an ONB on the joint system AB.

Embedding Linear Operators

We embed the bounded linear operators $\mathscr{L}(A)$ into $\mathscr{L}(AB)$ by taking a tensor product with the identity on B. We often omit to write this identity explicitly and instead use subscripts to indicate on which system an operator acts. For example, for any $L_A \in \mathscr{L}(A)$ and $|v\rangle_{AB} \in \mathscr{H}_{AB}$ as in (2.26), we write

$$L_A\, |v\rangle_{AB} = L_A \otimes I_B\, |v\rangle_{AB} = \sum_x \lambda_x\, L_A\, |e_x\rangle_A \otimes |f_x\rangle_B \tag{2.27}$$

Clearly, $\|L_A \otimes I_B\| = \|L_A\|$, and in fact, more generally for all $L_A \in \mathscr{L}(A)$ and $L_B \in \mathscr{L}(B)$, we have

$$\|L_A \otimes L_B\| = \|L_A\| \cdot \|L_B\|. \tag{2.28}$$

We say that two operators $K, L \in \mathscr{L}(A)$ *commute* if $[K, L] := KL - LK = 0$. Clearly, elements of $\mathscr{L}(A)$ and $\mathscr{L}(B)$ mutually commute as operators in $\mathscr{L}(AB)$, i.e. for all $L_A \in \mathscr{L}(A)$, $K_B \in \mathscr{L}(B)$, we have $[L_A \otimes I_B, I_A \otimes K_B] = 0$.

Finally, every linear operator $L_{AB} \in \mathscr{L}(AB)$ has a decomposition

$$L_{AB} = \sum_k L_A^k \otimes L_B^k, \quad \text{where} \quad L_A^k \in \mathscr{L}(A),\ L_B^k \in \mathscr{L}(B) \tag{2.29}$$

Similarly, every self-adjoint operator $L_{AB} \in \mathscr{L}^\dagger(AB)$ decomposes in the same way but now $L_A^k \in \mathscr{L}^\dagger(A)$ and $L_B^k \in \mathscr{L}^\dagger(B)$ can be chosen self-adjoint as well. However, crucially, it is not always possible to decompose a positive semi-definite operator into products of positive semi-definite operators in this way.

Representing Traces of Matrix Products Using Tensor Spaces

Let us next consider trace terms of the form $\mathrm{Tr}_A(K_A L_A)$ where $K_A, L_A \in \mathscr{L}(A)$ are general linear operators and $\mathscr{H}_A$ is finite-dimensional. It is often convenient to represent such traces as follows.

First, we introduce an auxiliary system A' such that $\mathscr{H}_A$ and $\mathscr{H}_{A'}$ are isomorphic (i.e. they have the same dimension). Furthermore, we fix a pair of bases $\{|e_x\rangle_A\}_x$ of A and $\{|e_x\rangle_{A'}\}_x$ of A'. (We can use the same index set here since these spaces are isomorphic.) Clearly every linear operator on A has a natural embedding into A' given by this isomorphism. Using these bases, we further define a rank one operator $\Psi \in \mathscr{S}(AA')$ in its Schmidt decomposition as

$$|\Psi\rangle_{AA'} = \sum_x |x\rangle_A \otimes |x\rangle_{A'}. \tag{2.30}$$

(Note that this state has norm $\|\Psi\|_* = d_A$, which is why this discussion is restricted to finite dimensions.) Using the matrix representation of the transpose in (2.11), we now observe that $L_A \otimes I_{A'} |\Psi\rangle_{AA'} = I_A \otimes L_{A'}^T |\Psi\rangle_{AA'}$ and, therefore,

$$\mathrm{Tr}(K_A L_A) = \langle\Psi| K_A L_A |\Psi\rangle = \langle\Psi|_{AA'} K_A \otimes L_{A'}^T |\Psi\rangle_{AA'}. \tag{2.31}$$

We will encounter this representation many times and keep Ψ thus reserved for this purpose, without going through the construction explicitly every time.[6]

Marginals of Functionals

Given a bipartite system AB that consists of two sets of operators $\mathscr{L}(A)$ and $\mathscr{L}(B)$, we now want to specify how a trace-class operator $\xi_{AB} \in \mathscr{T}(AB)$ acts on $\mathscr{L}(A)$. For any $L_A \in \mathscr{L}(A)$, we have

$$F_{\xi_{AB}}(L_A) = \langle \xi_{AB}, L_A \otimes I_B\rangle = \mathrm{Tr}\left(\xi_{AB}^\dagger\, L_A \otimes I_B\right) = \mathrm{Tr}_A\left(\mathrm{Tr}_B(\xi_{AB}^\dagger)\, L_A\right), \tag{2.32}$$

where we simply used that $\mathrm{Tr}_{AB}(\cdot) = \mathrm{Tr}_A(\mathrm{Tr}_B(\cdot))$ where Tr_B as defined in (2.16) naturally embeds as a map from $\mathscr{T}(AB)$ into $\mathscr{T}(A)$, i.e.

[6] Note that Ψ is an (unnormalized) maximally entangled state, usually denoted ψ.

$$\mathrm{Tr}_B\left(\xi_{AB}^{\dagger}\right) = \sum_x \left(\langle e_x|_A \otimes I_B\right)\xi_{AB}^{\dagger}\left(|e_x\rangle_A \otimes I_B\right) = \mathrm{Tr}_B\left(\xi_{AB}\right)^{\dagger}. \tag{2.33}$$

This is also called the *partial trace* and will be discussed further in the context of completely bounded maps in Sect. 2.6.2.

The above discussion allows us to define the *marginal* on A of the trace-class operator $\xi_{AB} \in \mathscr{T}(A)$ as follows:

$$\xi_A := \mathrm{Tr}_B\left(\xi_{AB}\right) \quad \text{such that} \quad F_{\xi_{AB}}(L_A) = F_{\xi_A}(L_A) = \langle \xi_A, L_A \rangle. \tag{2.34}$$

We usually do not introduce marginals explicitly. For example, if we introduce a trace-class operator ξ_{AB} then its marginals ξ_A and ξ_B are implicitly defined as well.

2.4.2 Separable States and Entanglement

The occurrence of entangled states on two or more quantum systems is one of the most intriguing features of the formalism of quantum mechanics.

We call a positive operator $M_{AB} \in \mathscr{P}(AB)$ of a joint quantum system AB **separable** if it can be written in the form

$$M_{AB} = \sum_{k \in \mathscr{K}} L_A(k) \otimes K_B(k), \quad \text{where} \quad L_A(k) \in \mathscr{P}(A),\ K_B(k) \in \mathscr{P}(B), \tag{2.35}$$

for some index set $\mathscr{K}$. Otherwise, it is called **entangled**.

The prime example of an entangled state is the *maximally entangled* state. For two quantum systems A and B of finite dimension, a maximally entangled state is a state of the form

$$|\psi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_x |e_x\rangle_A \otimes |f_x\rangle_B, \quad d = \min\{d_A, d_B\} \tag{2.36}$$

where $\{e_x\}_x$ is an ONB of A and $\{f_x\}_x$ is an ONB of B.

This state cannot be written in the form (2.35) as the following argument, due to Peres [131] and Horodecki [89], shows. Consider the operation $(\cdot)^{T_B}$ of taking a *partial transpose* on the system B with regards to $\{f_x\}_x$ on B. Applied to separable states of the from (2.35), this always results in a state, i.e.

$$\rho_{AB}^{T_B} = \sum_k \sigma_A(k) \otimes \big(\tau_B(k)\big)^{T_B} \geq 0. \tag{2.37}$$

Is positive semi-definite. Applied to ψ_{AB}, however, we get

$$\psi_{AB}^{T_B} = \frac{1}{d}\sum_{x,x'} |e_x\rangle\langle e_{x'}| \otimes \big(|f_x\rangle\langle f_{x'}|\big)^{T_B} = \frac{1}{d}\sum_{x,x'} |e_x\rangle\langle e_{x'}| \otimes |f_{x'}\rangle\langle f_x|. \tag{2.38}$$

This operator is not positive semi-definite. For example, we have

$$\langle\phi|\psi_{AB}^{T_B}|\phi\rangle = -\frac{2}{d}, \quad \text{where} \quad |\phi\rangle = |e_1\rangle \otimes |e_2\rangle - |e_2\rangle \otimes |e_1\rangle. \tag{2.39}$$

Generally, we have seen that a bipartite state is separable only if it remains positive semi-definite under the partial transpose. The converse is not true in general.

2.4.3 Purification

Consider any state $\rho_{AB} \in \mathscr{S}(AB)$, and its marginals ρ_A and ρ_B. Then we say that ρ_{AB} is an *extension* of ρ_A and ρ_B. Moreover, if ρ_{AB} is pure, we call it a *purification* of ρ_A and ρ_B. Moreover, we can always construct a purification of a given state $\rho_A \in \mathscr{S}(A)$. Let us say that ρ_A has eigenvalue decomposition

$$\rho_A = \sum_x \lambda_x\, |e_x\rangle\langle e_x|_A, \quad \text{then the state} \quad |\rho\rangle_{AA'} = \sum_x \sqrt{\lambda_x}\, |e_x\rangle_A \otimes |e_x\rangle_{A'} \tag{2.40}$$

is a *purification* of ρ_A. Here, A' is an auxiliary system of the same dimension as A and $\{|e_x\rangle_{A'}\}_x$ is any ONB of A'. Clearly, $\mathrm{Tr}_{A'}(\rho_{AA'}) = \rho_A$.

2.4.4 Classical-Quantum Systems

An important special case are joint systems where one part consists of a classical system. Events $M \in \mathscr{P}_\bullet(XA)$ on such joint systems can be decomposed as

$$M_{XA} = \sum_x |x\rangle\langle x|_X \otimes M_A(x) = \bigoplus_x M_A(x), \quad \text{where} \quad M_A(x) \in \mathscr{P}_\bullet(A). \tag{2.41}$$

Moreover, we call states of such systems *classical-quantum* states. For example, consistent with our notation for classical systems in (2.25), a state $\rho_{XA} \in \mathscr{S}_\bullet(XA)$ can be decomposed as

$$\rho_{XA} = \sum_x |x\rangle\langle x|_X \otimes \rho_A(x), \quad \text{where} \quad \rho_A(x) \geq 0, \quad \sum_x \mathrm{Tr}\left(\rho_A(x)\right) \leq 1. \tag{2.42}$$

Clearly, $\rho_A(x) \in \mathscr{S}_\bullet(A)$ is a sub-normalized density operator on A. Furthermore, comparing with (2.35), it is evident that such states are always separable.

If $\rho_{XA} \in \mathscr{S}_\circ(XA)$, it is sometimes more convenient to instead further decompose

$$\rho_A(x) = \rho(x)\hat{\rho}_A(x), \tag{2.43}$$

where $\rho(x)$ is a probability mass function and $\hat{\rho}_A(x) \in \mathscr{S}_\circ(A)$ normalized as well.

2.5 Functions on Positive Operators

Besides the inverse, we often need to lift other continuous real-valued functions to positive semi-definite operators. For any continuous function $f : \mathbb{R}_+ \setminus \{0\} \to \mathbb{R}$ and $M \in \mathscr{P}(A)$, we use the convention

$$f(M) = \sum_{x:\lambda_x \neq 0} f(\lambda_x)\, |e_x\rangle\langle e_x|. \tag{2.44}$$

If the resulting operator is bounded (e.g. if the spectrum of M is compact). That is, as for the generalized inverse, we simply ignore the kernel of M.[7] By definition, we thus have $f(UMU^\dagger) = Uf(M)U^\dagger$ for any unitary U. Moreover, we have

$$Lf(L^\dagger L) = f(LL^\dagger)L, \tag{2.45}$$

which can be verified using the *polar decomposition*, stating that we can always write $L = U|L|$ for some unitary operator U. An important example is the *logarithm*, defined as $\log M = \sum_{x:\lambda_x \neq 0} \log \lambda_x \;|e_x\rangle\langle e_x|$.

Let us in the following restrict our attention to the finite-dimensional case. Notably, trace functionals of the form $M \mapsto \mathrm{Tr}(f(M))$ inherit continuity, monotonicity, concavity and convexity from f (see, e.g., [34]). For example, for any monotonically increasing continuous function f, we have

$$\mathrm{Tr}(f(M)) \leq \mathrm{Tr}(f(N)) \qquad \text{for all} \quad M, N \in \mathscr{P}(A) \quad \text{with} \quad M \leq N. \tag{2.46}$$

[7] This convention is very useful to keep the presentation in the following chapters concise, but some care is required. If $\lim_{\varepsilon \to 0} f(\varepsilon) \neq 0$, then $M \mapsto f(M)$ is not necessarily continuous even if f is continuous on its support.

Table 2.2 Examples of operator monotone, concave and convex functions

Function	Range	Op. monotone	Op. anti-monotone	Op. convex	Op. concave
$\sqrt{t}$	$[0,\infty)$	Yes	No	No	Yes
t^2	$[0,\infty)$	No	No	Yes	No
$\frac{1}{t}$	$(0,\infty)$	No	Yes	Yes	No
t^α		$\alpha \in [0,1]$	$\alpha \in [-1,0)$	$\alpha \in [-1,0) \cup [1,2]$	$\alpha \in (0,1]$
$\log t$	$(0,\infty)$	Yes	No	No	Yes
$t \log t$	$[0,\infty)$	No	No	Yes	No

Note in particular that t^α is neither operator monotone, convex nor concave for $\alpha < -1$ and $\alpha > 2$

Operator Monotone and Concave Functions

Here we discuss classes of functions that, when lifted to positive semi-definite operators, retain their defining properties. A function $f : \mathbb{R}_+ \to \mathbb{R}$ is called *operator monotone* if

$$M \leq N \implies f(M) \leq f(N) \quad \text{for all} \quad M, N \geq 0. \tag{2.47}$$

If f is operator monotone then $-f$ is *operator anti-monotone*. Furthermore, f is called *operator convex* if

$$\lambda f(M) + (1-\lambda) f(N) \geq f\big(\lambda M + (1-\lambda)N\big) \quad \text{for all} \quad M, N \geq 0 \tag{2.48}$$

and $\lambda \in [0, 1]$. If this holds with the inequality reversed, then the function is called *operator concave*. These definitions naturally extend to functions $f : (0,\infty) \to \mathbb{R}$, where we consequently choose $M, N > 0$.

There exists a rich theory concerning such functions and their properties (see, for example, Bhatia's book [26]), but we will only mention a few prominent examples in Table 2.2 that will be of use later.

We say that a two-parameter function is *jointly concave* (*jointly convex*) if it is concave (convex) when we take convex combinations of input tuples. Lieb [106] and Ando [4] established the following extremely powerful result. The map

$$\mathscr{P}(A) \times \mathscr{P}(B) \to \mathscr{P}(AB), \quad (M_A, N_B) \mapsto f\big(M_A \otimes N_B^{-1}\big) M_A \otimes I_B \tag{2.49}$$

is jointly convex on strictly positive operators if $f : (0,\infty) \to \mathbb{R}$ is operator monotone. This is *Ando's convexity theorem* [4]. In particular, we find that the functional

$$(M_A, N_B) \mapsto \langle \Psi | K \cdot \big(M_A \otimes N_{B'}^{-T}\big)^{\alpha-1} M_A \cdot K^\dagger |\Psi\rangle_{BB'} = \mathrm{Tr}_A(M_A^\alpha K^\dagger N_B^{1-\alpha} K) \tag{2.50}$$

for any $K \in \mathscr{L}(A, B)$ is jointly concave for $\alpha \in (0, 1)$ and jointly convex for $\alpha \in (1, 2)$. The former is known as Lieb's concavity theorem. Since this will be used extensively, we include a derivation of this particular result in Appendix.

2.6 Quantum Channels

Quantum channels are used to model the time evolution of physical systems. There are two equivalent ways to model a quantum channel, and we will see that they are intimately related. In the Schrödinger picture, the events are fixed and the state of a system is time dependent. Consequently, we model evolutions as quantum channels acting on the space of density operators. In the Heisenberg picture, the observable events are time dependent and the state of a system is fixed, and we thus model evolutions as adjoint quantum channels acting on events.

2.6.1 Completely Bounded Maps

Here, we introduce linear maps between bounded linear operators on different systems, and their adjoints, which map between functionals on different systems. For later convenience, we use calligraphic letters to denote the latter maps, for example $\mathcal{E}$ and $\mathcal{F}$ and use the adjoint notation for maps between bounded linear operators. The action of a linear map on an operator in a tensor space is well-defined by linearity via the decomposition in (2.29), and as for linear operators, we usually omit to make this embedding explicit.

The set of *completely bounded* (CB) linear maps from $\mathscr{L}(A)$ to $\mathscr{L}(B)$ is denoted by $\mathrm{CB}(A, B)$. Completely bounded maps $\mathcal{E}^\dagger \in \mathrm{CB}(A, B)$ have the defining property that for any operator $L_{AC} \in \mathscr{L}(AC)$ and any auxiliary system C, we have $\|\mathcal{E}^\dagger(L_{AC})\| < \infty$.[8] We then define the linear map $\mathcal{E}$ from $\mathscr{T}(A)$ to $\mathscr{T}(B)$ as the *adjoint map* for some $\mathcal{E}^\dagger \in \mathrm{CB}(B, A)$ via the sesquilinear form. Namely, $\mathcal{E}$ is defined as the unique linear map satisfying

$$\langle \mathcal{E}(\xi), L \rangle = \langle \xi, \mathcal{E}^\dagger(L) \rangle \qquad \text{for all} \qquad \xi \in \mathscr{T}(A),\ L \in \mathscr{L}(B). \tag{2.51}$$

Clearly, $\mathcal{E}$ maps $\mathscr{T}(A)$ into $\mathscr{T}(B)$. Moreover, for any ξ_{AC} in $\mathscr{T}(AC)$, we have

$$\|\mathcal{E}(\xi_{AC})\|_* = \sup \left\{ \left| \left\langle \xi_{AC}, \mathcal{E}^\dagger(L_{BC}) \right\rangle \right| : L_{BC} \in \mathscr{L}_\bullet(BC) \right\} < \infty. \tag{2.52}$$

So these maps are in fact completely bounded in the trace norm and we collect them in the set $\mathrm{CB}_*(A, B)$. Again, in finite dimensions $\mathrm{CB}(A, B)$ and $\mathrm{CB}_*(A, B)$ coincide.

[8] It is noteworthy that the weaker condition that the map be bounded, i.e. $\|\mathcal{E}^\dagger(L_A)\| < \infty$, is not sufficient here and in particular does not imply that the map is completely bounded. In contrast, bounded linear operators in $\mathscr{L}(A)$ are in fact also completely bounded in the above sense.

2.6.2 Quantum Channels

Physical channels necessarily map positive functionals onto positive functionals. A map $\mathcal{E} \in \mathrm{CB}_*(A, B)$ is called *completely positive* (CP) if it maps $\mathscr{S}(AC)$ to $\mathscr{S}(BC)$ for any auxiliary system C, namely if

$$\langle \mathcal{E}(\omega_{AC}), M_{BC} \rangle \geq 0 \quad \text{for all} \quad \omega \in \mathscr{S}(AC),\ M \in \mathscr{P}(BC). \tag{2.53}$$

A map $\mathcal{E}$ is CP if and only if $\mathcal{E}^\dagger$ is CP, in the respective sense. The set of all CP maps from $\mathscr{T}(A)$ to $\mathscr{T}(B)$ is denoted $\mathrm{CP}(A, B)$.

Physical channels in the Schrödinger picture are modeled by completely positive trace-preserving maps, or quantum channels.

A **quantum channel** is a map $\mathcal{E} \in \mathrm{CP}(A, B)$ that is **trace-preserving**, namely a map that satisfies

$$\mathrm{Tr}(\mathcal{E}(\xi)) = \mathrm{Tr}(\xi) \quad \text{for all} \quad \xi \in \mathscr{T}(A). \tag{2.54}$$

Naturally, such maps take states to states, more precisely, they map $\mathscr{S}_\circ(A)$ to $\mathscr{S}_\circ(B)$ and $\mathscr{S}_\bullet(A)$ to $\mathscr{S}_\bullet(B)$. The corresponding adjoint quantum channel $\mathcal{E}^\dagger$ from $\mathscr{L}(B)$ to $\mathscr{L}(A)$ in the Heisenberg picture is a completely positive and *unital* map, namely it satisfies $\mathcal{E}^\dagger(I_A) = I_B$. In fact, a map $\mathcal{E}$ is trace-preserving if and only if $\mathcal{E}^\dagger$ is unital. Unital maps take $\mathscr{P}_\bullet(B)$ to $\mathscr{P}_\bullet(A)$ and thus map events to events. Clearly,

$$\Pr_{\mathcal{E}(\rho)}(M) = \langle \mathcal{E}(\rho), M \rangle = \left\langle \rho, \mathcal{E}^\dagger(M) \right\rangle = \Pr_\rho\left(\mathcal{E}^\dagger(M)\right). \tag{2.55}$$

Let us summarize some further notation:

- We denote the set of all completely positive trace-preserving (CPTP) maps from $\mathscr{T}(A)$ to $\mathscr{T}(B)$ by $\mathrm{CPTP}(A, B)$.
- The set of all CP unital maps from $\mathscr{L}(A)$ to $\mathscr{L}(B)$ is denoted $\mathrm{CPU}(A, B)$.
- Finally, a map $\mathcal{E} \in \mathrm{CP}(A, B)$ is called *trace-non-increasing* if $\mathrm{Tr}(\mathcal{E}(\omega)) \leq \mathrm{Tr}(\omega)$ for all $\omega \in \mathscr{S}(A)$. A CP map is trace-non-increasing if and only if its adjoint is *sub-unital*, i.e. it satisfies $\mathcal{E}^\dagger(I_B) \leq I_A$.

Some Examples of Channels

The simplest example of such a CP map is the *conjugation* with an operator $L \in \mathscr{L}(A, B)$, that is the map $\mathcal{L} : \xi \mapsto L\xi L^\dagger$. We will often use the following basic property of completely positive maps. Let $\mathcal{E} \in \mathrm{CP}(A, B)$, then

$$\xi \geq \zeta \implies \mathcal{E}(\xi) \geq \mathcal{E}(\zeta) \quad \text{for all} \quad \xi, \zeta \in \mathscr{T}(A). \tag{2.56}$$

As a consequence, we take note of the following property of positive semi-definite operators. For any $M \in \mathscr{P}(A)$, $\xi \in \mathscr{S}(A)$, we have

$$\mathrm{Tr}(\xi M) = \mathrm{Tr}\left(\sqrt{M}\xi\sqrt{M}\right) \geq 0, \tag{2.57}$$

where the last inequality follows from the fact that the conjugation with $\sqrt{M}$ is a completely positive map. In particular, if $L, K \in \mathscr{L}(A)$ satisfy $L \geq K$, we find $\mathrm{Tr}(\xi L) \geq \mathrm{Tr}(\xi K)$.

An instructive example is the embedding map $L_A \mapsto L_A \otimes I_B$, which is completely bounded, CP and unital. Its adjoint map is the CPTP map Tr_B, the partial trace, as we have seen in Sect. 2.4.1. Finally, for a POVM $x \mapsto M_A(x)$, we consider the measurement map $\mathcal{M} \in \mathrm{CPTP}(A, X)$ given by

$$\mathcal{M} : \rho_A \mapsto \sum_x |x\rangle\langle x| \ \mathrm{Tr}(\rho_A M_A(x)). \tag{2.58}$$

This maps a quantum system into a classical system with a state corresponding to the probability mass function $\rho(x) = \mathrm{Tr}(\rho_A M_A(x))$ that arises from Born's rule. If the events $\{M_A(x)\}_x$ are rank-one projectors, then this map is also unital.

2.6.3 Pinching and Dephasing Channels

Pinching maps (or channels) constitute a particularly important class of quantum channels that we will use extensively in our technical derivations. A *pinching map* is a channel of the form $\mathcal{P} : L \mapsto \sum_x P_x \, L \, P_x$ where $\{P_x\}_x$, $x \in [m]$ are orthogonal projectors that sum up to the identity. Such maps are CPTP, unital and equal to their own adjoints. Alternatively, we can see them as *dephasing* operations that remove off-diagonal blocks of a matrix. They have two equivalent representations:

$$\mathcal{P}(L) = \sum_{x\in[m]} P_x L P_x = \frac{1}{m}\sum_{y\in[m]} U_y L U_y^\dagger, \quad \text{where} \quad U_y = \sum_{x\in[m]} e^{\frac{2\pi i y x}{m}} P_x \tag{2.59}$$

are unitary operators. Note also that $U_m = I$.

For any self-adjoint operator $H \in \mathscr{L}^\dagger(A)$ with eigenvalue decomposition $H = \sum_x \lambda_x \, |e_x\rangle\langle e_x|$, we define the set $\mathrm{spec}(H) = \{\lambda_x\}_x$ and its cardinality, $|\,\mathrm{spec}(H)|$, is the number of distinct eigenvalues of H. For each $\lambda \in \mathrm{spec}(H)$, we also define $P_\lambda = \sum_{x:\lambda_x=\lambda} |e_x\rangle\langle e_x|$ such that $H = \sum_\lambda \lambda P_\lambda$ is its *spectral decomposition*. Then, the *pinching map* for this spectral decomposition is denoted

$$\mathcal{P}_H : L \mapsto \sum_{\lambda\in\mathrm{spec}(H)} P_\lambda \, L \, P_\lambda. \tag{2.60}$$

Clearly, $\mathcal{P}_H(H) = H$, $\mathcal{P}_H(L)$ commutes with H, and $\mathrm{Tr}(\mathcal{P}_H(L)H) = \mathrm{Tr}(LH)$.

For any $M \in \mathscr{P}(A)$, using the second expression in (2.59) and the fact that $U_x M U_x^\dagger \geq 0$, we immediately arrive at

$$\mathcal{P}_H(M) = \frac{1}{|\operatorname{spec}(H)|} \sum_{y\in[m]} U_y M U_y^\dagger \geq \frac{1}{|\operatorname{spec}(H)|} M. \tag{2.61}$$

This is Hayashi's *pinching inequality* [74].

Finally, if f is operator concave, then for every pinching $\mathcal{P}$, we have

$$f(\mathcal{P}(M)) = f\left(\frac{1}{m}\sum_{x\in[m]} U_x M U_x^\dagger\right) \geq \frac{1}{m}\sum_{x\in[m]} f\big(U_x M U_x^\dagger\big) \tag{2.62}$$

$$= \frac{1}{m}\sum_{x\in[m]} U_x f(M) U_x^\dagger = \mathcal{P}(f(M)). \tag{2.63}$$

This is a special case of the *operator Jensen inequality* established by Hansen and Pedersen [71]. For all $H \in \mathscr{L}^\dagger(A)$, every operator concave function f defined on the spectrum of H, and all unital maps $\mathcal{E} \in \mathrm{CPU}(A, B)$, we have

$$f(\mathcal{E}(H)) \geq \mathcal{E}(f(H)). \tag{2.64}$$

2.6.4 Channel Representations

The following representations for trace non-increasing and trace preserving CP maps are of crucial importance in quantum information theory.

Kraus Operators

Every CP map can be represented as a sum of conjugations of the input [82, 83]. More precisely, $\mathcal{E} \in \mathrm{CP}(A, B)$ if and only if there exists a set of linear operators $\{E_k\}_k$, $E_k \in \mathscr{L}(A, B)$ such that

$$\mathcal{E}(\xi) = \sum_k E_k \xi\, E_k^\dagger \quad \text{for all} \quad \xi \in \mathscr{T}(A). \tag{2.65}$$

Furthermore, such a channel is trace-preserving if and only if $\sum_k E_k^\dagger E_k = I$, and trace-non-increasing if and only if $\sum_k E_k^\dagger E_k \leq I$. The operators $\{E_k\}$ are called *Kraus operators*. Moreover, the adjoint $\mathcal{E}^\dagger$ of $\mathcal{E}$ is completely positive and has Kraus operators $\{E_k^\dagger\}$ since

$$\mathrm{Tr}\left(\xi\,\mathcal{E}^\dagger(L)\right) = \mathrm{Tr}\left(\mathcal{E}(\xi)L\right) = \mathrm{Tr}\left(\xi \sum_k E_k^\dagger L\, E_k\right). \tag{2.66}$$

Stinespring Dilation

Moreover, every CP map can be decomposed into its *Stinespring dilation* [147]. That is, $\mathcal{E} \in \mathrm{CP}(A, B)$ if and only if there exists a system C and an operator $L \in \mathscr{L}(A, BC)$ such that

$$\mathcal{E}(\xi) = \mathrm{Tr}_C(L\xi L^\dagger) \quad \text{for all} \quad \xi \in \mathscr{T}(A). \tag{2.67}$$

Moreover, if $\mathcal{E}$ is trace-preserving then $L = U$, where $U \in \mathscr{L}_\bullet(A, BC)$ is an isometry. If $\mathcal{E}$ is trace-non-increasing, then $L = PU$ is an isometry followed by a projection $P \in \mathscr{P}_\bullet(C)$.

Choi-Jamiolkowski Isomorphism

For finite-dimensional Hilbert spaces, the *Choi-Jamiolkowski isomorphism* [96] between bounded linear maps from A to B and linear functionals on $A'B$ is given by

$$\Gamma : \mathscr{T}(\mathscr{T}(A), \mathscr{T}(B)) \to \mathscr{T}(A'B), \quad \mathcal{E} \mapsto \gamma_{A'B}^{\mathcal{E}} = \mathcal{E}\big(|\Psi\rangle\langle\Psi|_{A'A}\big), \tag{2.68}$$

where the state $\gamma_{A'B}^{\mathcal{E}}$ is called the Choi-Jamiolkowski state of $\mathcal{E}$. The inverse operation, Γ^{-1}, maps linear functionals to bounded linear maps

$$\Gamma^{-1} : \gamma_{A'B} \mapsto \Big\{\mathcal{E}^\gamma : \rho_A \mapsto \mathrm{Tr}_{A'}\left(\gamma_{A'B}(I_B \otimes \rho_{A'}^T)\right)\Big\}, \tag{2.69}$$

where the transpose is taken with regards to the Schmidt basis of Ψ.

There are various relations between properties of bounded linear maps and properties of the corresponding Choi-Jamiolkowski functionals, for example:

$$\mathcal{E} \text{ is completely positive} \Leftrightarrow \gamma_{A'B}^{\mathcal{E}} \geq 0 \tag{2.70}$$

$$\mathcal{E} \text{ is trace-preserving} \Leftrightarrow \mathrm{Tr}_B(\gamma_{A'B}^{\mathcal{E}}) = I_{A'} \tag{2.71}$$

$$\mathcal{E} \text{ is unital} \Leftrightarrow \mathrm{Tr}_{A'}(\gamma_{A'B}^{\mathcal{E}}) = I_B. \tag{2.72}$$

2.7 Background and Further Reading

Nielsen and Chuang's book [125] offers a good introduction to the quantum formalism. Hayashi's [75] and Wilde's [174] books both also carefully treat the concepts relevant for quantum information theory in finite dimensions. Finally, Holevo's recent book [88] offers a comprehensive mathematical introduction to quantum information processing in finite and infinite dimensions.

Operator monotone functions and other aspects of matrix analysis are covered in Bhatia's books [26, 27], and Hiai and Petz' book [87].

http://www.springer.com/978-3-319-21890-8

Quantum Information Processing with Finite Resources
Mathematical Foundations
Tomamichel, M.
2016, IX, 138 p. 4 illus. in color., Softcover
ISBN: 978-3-319-21890-8