Basics of finite-dimensional quantum information theory

Roy Araiza and Felix Leditzky

Department of Mathematics & IQUIST, University of Illinois at Urbana-Champaign

{raraiza,leditzky}@illinois.edu

Spring Term 2022

Contents

1 Tensor products

Our goal for this section will be to introduce the "Universal mapping property" for algebraic tensor products (the vector space tensor product). In short, the algebraic tensor of vector spaces is the universal object which linearizes bilinear maps on the cartesian product of said vector spaces. Thus, given vector spaces $\mathcal{X}, \mathcal{Y}$, and any bilinear map $\varphi : \mathcal{X} \times \mathcal{Y} \to \mathbb{C}$, there exists a linear map $U : \mathcal{X} \otimes \mathcal{Y} \to \mathbb{C}$ such that $\varphi = U \otimes$. This notation will be defined shortly. For this section we will adopt the following notation: the space of all bilinear maps from $\mathcal{X} \times \mathcal{Y} \to \mathbb{C}$ will be denoted $B(\mathcal{X}, \mathcal{Y})$, and we will at times reference such

maps as *bilinear forms*. The space of all linear maps between two vector spaces $\mathcal{X}, \mathcal{Y}$, will be denoted $\mathcal{L}(\mathcal{X} : \mathcal{Y})$. Given a vector space $\mathcal{X}$, we define the *vector space dual*, $\mathcal{X}'$, as the space $\mathcal{L}(\mathcal{X} : \mathbb{C})$. Note that $\langle \mathcal{X}, \mathcal{X}' \rangle$ form a *dual pair* in the sense that there exists a bilinear map $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X}' \to \mathbb{C}$ which satisfies the following two properties: (i) given $x \in \mathcal{X}$, then if $\langle x, x' \rangle = 0$ for all $x' \in \mathcal{X}'$ then $x = 0$. (ii) given $x' \in \mathcal{X}'$ then if $\langle x, x' \rangle = 0$ for all $x \in \mathcal{X}$ then $x' = 0$. Note by defining said dual pair, we are also considering the vector space $\mathcal{X}$ as a vector subspace of its *bidual*, $\mathcal{X}''$. Here, the bidual is defined as the vector space dual of $\mathcal{X}'$. In particular, given any $x \in \mathcal{X}$ then we define an element $\hat{x} \in \mathcal{L}(\mathcal{X}' : \mathbb{C})$ as $\hat{x}(x') := \langle x, x' \rangle$. The mapping $x \mapsto \hat{x}$ is a linear embedding of vector spaces, and therefore one may always realize the vector space $\mathcal{X}$ as a vector subspace of its bidual.

Let $x \in \mathcal{X}, y \in \mathcal{Y}$ and define the map $x \otimes y \in B(\mathcal{X}, \mathcal{Y} : \mathbb{C})'$ by $x \otimes y(\varphi) := \varphi(x, y)$. By bilinearity for any such φ, it is automatic that such a map is linear. We define the *algebraic tensor product*, $\mathcal{X} \otimes \mathcal{Y}$, as the space

$$\mathcal{X} \otimes \mathcal{Y} := \operatorname{span}\{x \otimes y : x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{1}$$

It is readily checked the $\otimes$ is linear in both variables. This is to say,

1. $x_1 + x_2 \otimes y = x_1 \otimes y + x_2 \otimes y$.
2. $x \otimes y_1 + y_2 = x \otimes y_1 + x \otimes y_2$.
3. $\lambda(x \otimes y) = \lambda x \otimes y = x \otimes \lambda y$.
4. $x \otimes 0 = 0 \otimes y = 0_{\mathcal{X} \otimes \mathcal{Y}}$.

These properties hold since such elements $u, v \in \mathcal{X} \otimes \mathcal{Y}$ satisfy $u = v$ if and only if $u(\varphi) = v(\varphi)$ for all $\varphi \in B(\mathcal{X}, \mathcal{Y} : \mathcal{Z})$. Another consequence of this is that such a decomposition

$$u = \sum_{i \leq N} \lambda_i x_i \otimes y_i,$$

is far from unique.

There are very nice properties which the algebraic tensor product satisfies. Given two linear maps $u \in L(\mathcal{X} : \tilde{\mathcal{X}})$ and $v \in L(\mathcal{Y} : \tilde{\mathcal{Y}})$, then they induce a linear map

$$u \otimes v \colon \mathcal{X} \otimes \mathcal{Y} \to \tilde{\mathcal{X}} \otimes \tilde{\mathcal{Y}},$$

defined on simple tensors as

$$u \otimes v(x \otimes y) := u(x) \otimes v(y),$$

and then extending linearly. Thus,

$$\mathcal{L}(\mathcal{X} : \tilde{\mathcal{X}}) \otimes \mathcal{L}(\mathcal{Y} : \tilde{\mathcal{Y}}) \subset \mathcal{L}(\mathcal{X} \otimes \mathcal{Y} : \tilde{\mathcal{X}} \otimes \tilde{\mathcal{Y}}).$$

Proposition 1.1. *Let $E \subset \mathcal{X}$ and $F \subset \mathcal{Y}$ be two linearly independent sets. Then the set*

$$E \otimes F := \{e \otimes f : e \in E, f \in F\},$$

is linearly independent in $\mathcal{X} \otimes \mathcal{Y}$.

Proof. Let I, J be finite sets and consider a linear combination $\sum_{i\in I, j\in J} \lambda_{ij} e_i \otimes f_j$. We wish to show $\sum_{i\in I, j\in J} \lambda_{ij} e_i \otimes f_j = 0$ if and only if $\lambda_{ij} = 0$ for each $i \in I, j \in J$. By definition, it follows $\sum_{i\in I, j\in J} \lambda_{ij} e_i \otimes f_j = 0$ if and only if $\sum_{i\in I, j\in J} \lambda_{ij} e_i \otimes f_j(\varphi) = 0$ for every $\varphi \in B(\mathcal{X}, \mathcal{Y} : \mathbb{C})$. In particular, $\sum_{i\in I, j\in J} \lambda_{ij} \varphi(e_i, f_j) = 0$ for every $\varphi \in B(\mathcal{X}, \mathcal{Y} : \mathcal{C})$. given $x \in \mathcal{X}', y' \in \mathcal{Y}'$, we consider bilinear forms of the form $\varphi_{(x',y')} : \mathcal{X} \times \mathcal{Y} \to \mathbb{C}$ defined as $\varphi_{(x',y')}(x, y) := \langle x, x'\rangle\langle y, y'\rangle$. By assumption, it must follow

$$\sum_{i\in I, j\in J} \lambda_{ij} e_i \otimes f_j \varphi_{(x',y')} = \sum_{i\in I, j\in J} \lambda_{ij} \varphi_{(x',y')}(e_i, f_j) = \sum_{i\in I, j\in J} \lambda_{ij} \langle e_i, x'\rangle\langle f_j, y'\rangle = x'(\sum_{i\in I, j\in J} \lambda_{ij} \langle f_j, y'\rangle e_i) = 0,$$

and this holds for all $x' \in \mathcal{X}'$. Since $\langle \mathcal{X}, \mathcal{X}'\rangle$ is a dual pair, it must follow

$$\sum_{i\in I, j\in J} \lambda_{ij} \langle f_j, y'\rangle e_i = 0.$$

But by linear independence of the set E, it must follow

$$\sum_{j\in J} \lambda_{ij} \langle f_j, y'\rangle = 0, \text{for each } i \in I.$$

This in turn implies $y'(\sum_{j\in J} \lambda_{ij} f_j) = 0$ for each $y' \in \mathcal{Y}'$, and for each $i \in I$. Since $\langle \mathcal{Y}, \mathcal{Y}'\rangle$ form a dual pair, it follows $\sum_{j\in J} \lambda_{ij} f_j = 0$ for each $i \in I$. By linear independence of F we therefore have $\lambda_{ij} = 0$ for each $j \in J$, and $i \in I$. This finishes the proof. □

Let $\varphi \in B(\mathcal{X}, \mathcal{Y} : \mathbb{C})$ and define the mapping $U_\varphi : \mathcal{X} \otimes \mathcal{Y} \to \mathbb{C}$ on simple tensors by $U_\varphi(x \otimes y) := \varphi(x, y)$ and then extend linearly. By bilinearity of φ, it is immediate that U_φ is linear. Furthermore, it follows $\varphi = U_\varphi \otimes$, i.e., the bilinear form φ factors through the algebraic tensor product $\mathcal{X} \otimes \mathcal{Y}$. Similarly, given any linear map $U : \mathcal{X} \otimes \mathcal{Y} \to \mathbb{C}$, then we may define the map $\varphi_U : \mathcal{X} \times \mathcal{Y} \to \mathbb{C}$ by $\varphi_U(x, y) := U(x \otimes y)$. By the very nature of the tensor product, it follows $\varphi_U \in B(\mathcal{X}, \mathcal{Y} : \mathbb{C})$, and $\varphi_U = U \otimes$. Furthermore,

$$\varphi_{U_\varphi} = \varphi, \quad \text{and} \quad U_{\varphi_U} = U.$$

Theorem 1.2 (The Universal Mapping Property)**.** *Let $\mathcal{X}$, and $\mathcal{Y}$ be vector spaces and suppose there exists a vector space $\mathcal{W}$ and a bilinear map $\tau : \mathcal{X} \times \mathcal{Y} \to \mathcal{W}$ such that given any bilinear map $\varphi : \mathcal{X} \times \mathcal{Y} \to \mathbb{C}$ there exists a linear map $U : \mathcal{W} \to \mathbb{C}$ such that $\varphi = U\tau$. Then there exists a linear isomorphism $\Psi : \mathcal{X} \otimes \mathcal{Y} \to \mathcal{W}$ such that $\tau(x, y) = \Psi(x \otimes y)$.*

The Universal Mapping property thus states that the algebraic tensor product of two vector spaces is the unique vector space, up to isomorphism, which linearizes bilinear maps. We point out that we could have replaced the scalar field $\mathbb{C}$ with a vector space $\mathcal{Z}$, and proceeded in an analogous fashion.

1.1 Hilbertian tensor products

We now wish to consider tensor products of Hilbert spaces. That is, tensor products of complete inner product spaces. Though the manuscript deals with finite-dimensional Hilbert spaces (and thus, isometric copies of $\mathbb{C}^n$ for some $n \in \mathbb{N}$), we will be stating some of the definitions and results in full generality.

Definition 1.3. A *Hilbert space* is defined to be a pair $(H, \langle\cdot|\cdot\rangle)$ where H is a vector space, $\langle\cdot|\cdot\rangle : H \times H \to \mathbb{C}$ is an inner product, and H is complete in the norm $\||\xi\rangle\|_H := \sqrt{\langle\xi|\xi\rangle}$.

In contrast to the mathematical convention, we will assume that inner products on Hilbert spaces are conjugate linear in their first coordinate, and linear in the second. Given two Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$, we wish to put a Hilbert space structure on the vector space $\mathcal{H}_A \otimes \mathcal{H}_B$. To this end, consider the following mapping

$$(\cdot|\cdot)_{\mathcal{H}_A\otimes\mathcal{H}_B} : \mathcal{H}_A \otimes \mathcal{H}_B \times \mathcal{H}_A \otimes \mathcal{H}_B \to \mathbb{C}, (|\psi_1\rangle_A \otimes |\psi_2\rangle_B, |\eta_1\rangle_A \otimes |\eta_2\rangle_B) \mapsto \langle\psi_1|\eta_1\rangle_A \langle\psi_2|\eta_2\rangle_B. \quad (2)$$

Extending linearly, this mapping defines an inner product on the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$. Throughout the manuscript, we may write the tensor product of two vectors as $|\psi\rangle \otimes |\eta\rangle$ or $|\psi\eta\rangle$. In the last section we proved that the tensor product of two linearly independent sets is linearly independent in the algebraic tensor product. We can in fact say more:

Proposition 1.4. *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be two Hilbert-spaces with respective orthonormal bases $\{e_i\}_{i\in I} \subset \mathcal{H}_A$ and $\{f_j\}_{j\in J} \subset \mathcal{H}_B$. Then the set*

$$\{e_i \otimes f_j : i \in I, j \in J\},$$

is an orthonormal basis for the inner product space $(\mathcal{H}_A \otimes \mathcal{H}_B, (\cdot|\cdot)_{\mathcal{H}_A\otimes\mathcal{H}_B})$.

We define the *Hilbert space tensor product* to be the completion of $\mathcal{H}_A \otimes \mathcal{H}_B$ with respect to the norm induced by the inner product $(\cdot|\cdot)_{\mathcal{H}_A\otimes\mathcal{H}_B}$. For the purposes of this manuscript, since all Hilbert spaces are already finite-dimensional, they are already complete. Thus, in the finite-dimensional case, the pair $(\mathcal{H}_A \otimes \mathcal{H}_B, (\cdot|\cdot)_{\mathcal{H}_A\otimes\mathcal{H}_B})$ is a finite-dimensional inner product space, and therefore is already complete. Recall that given a Hilbert space $\mathcal{H}$, then a linear operator $T \in \mathcal{L}(\mathcal{H})$ is *continuous* if and only if it is *bounded*. By once again restricting ourselves to finite-dimensional Hilbert spaces, one can show that any linear operator $T : \mathcal{H} \to \mathcal{H}$ is automatically continuous (thanks to finite-dimensionality). In particular, analogous with the algebraic tensor product, if $\mathcal{H}_A$ and $\mathcal{H}_B$ are two Hilbert spaces, then we have the following inclusion of vector spaces

$$\mathcal{L}(\mathcal{H}_A) \otimes \mathcal{L}(\mathcal{H}_B) \subset \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B),$$

with the simple tensors of linear operators defined on simple tensors as $N \otimes M(|\psi\eta\rangle) := N|\psi\rangle \otimes M|\eta\rangle$.

We have restricted our attention to bipartite systems. This is to say the tensor product of two Hilbert spaces, but our notions all extend to tensor products of more than two objects. For example, if $\mathcal{H}$ is a finite-dimensional Hilbert space, then we may wish to consider the Hilbert space

$$\bigotimes_{i=1}^{k} \mathcal{H}_i, \mathcal{H}_i = \mathcal{H}\,, \forall i \in \{1, \cdots, k\}.$$

It is also common to write the k^{th} iterated tensor product of a Hilbert space $\mathcal{H}$, with itself, as $\mathcal{H}^{\otimes k}$. An important instance will be if we wish consider linear operators on *k-qubits*. This is to say, linear operators $T \in \mathcal{L}((\mathbb{C}^2)^{\otimes k})$. We refer the reader to Section 2 for more details on qubits.

1.2 The partial trace and partial transpose

Consider a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_b$ and let $\{|i_A i_B\rangle\}$ be an orthonormal basis. Then we define the *partial trace* on the system $\mathcal{H}_A$ as the linear operator

$$\mathrm{tr}_A := \mathrm{tr} \otimes \mathrm{id}_B : \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_B),$$

defined on simple tensors by

$$\mathrm{tr}_A(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) = \mathrm{tr}(|a_1\rangle\langle a_2|)|b_1\rangle\langle b_2|.$$

In particular, if $\rho_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ then

$$\mathrm{tr}_A(\rho_{AB}) = \left(\sum_i \langle i_A| \otimes \mathbb{1}_B\right) \rho_{AB} \left(\sum_i |i_A\rangle \otimes \mathbb{1}_B\right) =: \rho_B. \tag{3}$$

Similarly, the partial trace of the system B is defined as

$$\mathrm{tr}_B := \mathrm{id}_A \otimes \mathrm{tr} \colon \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_A), \tag{4}$$

where if $\rho_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ then

$$\mathrm{tr}_B(\rho_{AB}) = \left(\sum_i \mathbb{1}_A \otimes \langle i_B|\right) \rho_{AB} \left(\sum_i \mathbb{1}_A \otimes |i_B\rangle\right) =: \rho_A. \tag{5}$$

We point out that as is the case with the trace $\mathrm{tr} \in \mathcal{L}(\cdot)$, the partial trace is basis independent.
To see the partial trace in action, let us look at our old friend, the EPR state (see Section 2),

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

$|\Phi^+\rangle$ is a vector on two qubits $\mathbb{C}^2 \otimes \mathbb{C}^2$, and in particular, $\Phi^+ := |\Phi^+\rangle\langle\Phi^+| \in \mathcal{L}(\mathbb{C}^2 \otimes \mathbb{C}^2)$. Let us trace out the second qubit and see what happens. First note

$$\Phi^+ = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|).$$

It then follows

$$\mathrm{tr}_{\mathbb{C}^2}(\psi) = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{1}{2}\mathbb{1}_{\mathbb{C}^2}.$$

In a similar fashion, one may extend the partial trace to the multipartite setting.
Recall that the *transposition* of an operator $\rho = \sum \alpha_{ij}|i\rangle\langle j|$ is defined as $T(\rho) = \sum \alpha_{ij}|j\rangle\langle i|$. We may then consider the *partial transpose* on a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$. The *partial transpose* on A is defined as the operator

$$T_A := T \otimes \mathrm{id}_B \colon \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B). \tag{6}$$

Similarly, the partial transpose on the system $\mathcal{H}_B$ is defined as the operator

$$T_B := \mathrm{id}_A \otimes T \colon \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \tag{7}$$

A helpful way to understand the partial transpose is via the block-matrix operator formulation. To this end, consider an operator $\rho \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ where we fix an orthonormal basis $\{i_A\} \subset \mathcal{H}_A$ and $\{i_B\} \subset \mathcal{H}_B$, and suppose $\dim \mathcal{H}_A = p$, and $\dim \mathcal{H}_B = q$. We may then write ρ as

$$\rho = (\rho_{i_A j_A})_{i_A, j_A = 1}^{p}$$

where

$$\rho_{i_A j_A} = \begin{pmatrix} \langle i_A \otimes 1_B|\rho|j_A \otimes 1_B\rangle & \langle i_A \otimes 1_B|\rho|j_A \otimes 2_B\rangle & \cdots & \langle i_A \otimes 1_B|\rho|j_A \otimes q\rangle \\ \langle i_A \otimes 2_B|\rho|j_A \otimes 1_B\rangle & \langle i_A \otimes 2_B|\rho|j_A \otimes 2_B\rangle & \cdots & \langle i_A \otimes 2_B|\rho|j_A \otimes q\rangle \\ \vdots & \ddots & \cdots & \vdots \\ \langle i_A \otimes q_B|\rho|j_A \otimes 1_B\rangle & \langle i_A \otimes q_B|\rho|j_A \otimes 2_B\rangle & \cdots & \langle i_A \otimes q_B|\rho|j_A \otimes q\rangle \end{pmatrix}$$

With this picture in mind, we then are able to immediately see how the partial transpose acts on each system. For notational convenience, let us simply write $\rho_{i_A j_A}$ as ρ_{ij}. We see,

$$T_A(\rho) = (\rho_{ji})_{i,j=1}^{p}, \text{ and } \quad T_B(\rho) = (\rho_{ij}^T)_{i,j=1}^{p},$$

$$\rho_{ij}^T = \begin{pmatrix} \langle i_A \otimes 1_B|\rho|j_A \otimes 1_B\rangle & \langle i_A \otimes 1_B|\rho|j_A \otimes 2_B\rangle & \cdots & \langle i_A \otimes 1_B|\rho|j_A \otimes q\rangle \\ \langle i_A \otimes 2_B|\rho|j_A \otimes 1_B\rangle & \langle i_A \otimes 2_B|\rho|j_A \otimes 2_B\rangle & \cdots & \langle i_A \otimes 2_B|\rho|j_A \otimes q\rangle \\ \vdots & \ddots & \cdots & \vdots \\ \langle i_A \otimes q|\rho|j_A \otimes 1_B\rangle & \langle i_A \otimes q|\rho|j_A \otimes 2_B\rangle & \cdots & \langle i_A \otimes q|\rho|j_A \otimes q\rangle \end{pmatrix}^T$$

$$= \begin{pmatrix} \langle i_A \otimes 1_B|\rho|j_A \otimes 1_B\rangle & \langle i_A \otimes 2_B|\rho|j_A \otimes 1_B\rangle & \cdots & \langle i_A \otimes q|\rho|j_A \otimes 1_B\rangle \\ \langle i_A \otimes 1_B|\rho|j_A \otimes 2_B\rangle & \langle i_A \otimes 2_B|\rho|j_A \otimes 2_B\rangle & \cdots & \langle i_A \otimes q|\rho|j_A \otimes 2_B\rangle \\ \vdots & \ddots & \cdots & \vdots \\ \langle i_A \otimes 1_B|\rho|j_A \otimes q\rangle & \langle i_A \otimes 2_B|\rho|j_A \otimes q\rangle & \cdots & \langle i_A \otimes q|\rho|j_A \otimes q_B\rangle \end{pmatrix}$$

In this spirit, suppose we once again consider the pure state $\Phi^+ \in \mathcal{L}(\mathbb{C}^2 \otimes \mathbb{C}^2)$ as above, and let us take the partial transpose on the second qubit. If we write the pure state Φ^+ in its block-matrix form we obtain

$$\Phi^+ = \frac{1}{2}\begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix}.$$

It then follows by taking the partial transpose on the second qubit we obtain

$$T_{\mathbb{C}^2}(\Phi^+) = \frac{1}{2}\begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix}.$$

We will revisit the above pure state in Section 2.3.

2 Quantum systems and quantum states

2.1 Pure and mixed quantum states

A quantum system is a (typically small, e.g., atomic scale and below) physical system with quantum degrees of freedom. The quantum system is called finite-dimensional if there is a finite (and hence discrete) number of degrees of freedom. Well-known examples are the polarization (horizontal or vertical)

of a photon; the spin (up or down) of an electron; the energy levels (ground state, excited states) of an atom.

Mathematically, a finite-dimensional quantum system is modeled by a finite-dimensional Hilbert space, typically denoted $\mathcal{H}$. Unless specified otherwise, we set $d = \dim \mathcal{H}$. In finite dimensions, a Hilbert space is nothing but a complex inner product space. Following Dirac's bra-ket notation, we will denote vectors in $\mathcal{H}$ by $|\psi\rangle$, and the inner product of vectors $|\phi\rangle, |\psi\rangle \in \mathcal{H}$ by $\langle\phi|\psi\rangle \in \mathbb{C}$. Here, $\langle\phi| = (|\phi\rangle)^\dagger$ is the adjoint of the vector $|\phi\rangle$, i.e., the linear functional

$$\langle\phi| \colon \mathcal{H} \to \mathbb{C}, |\psi\rangle \longmapsto \langle\phi|\psi\rangle. \tag{8}$$

In coordinates, $|\phi\rangle$ is a column vector, and the row vector $\langle\phi|$ is its conjugate transpose.

Definition 2.1 (Pure quantum states)**.** A *pure quantum state* $|\psi\rangle$ of a quantum system modeled by a Hilbert space $\mathcal{H}$ is a normalized vector in $\mathcal{H}$, i.e., $|\psi\rangle \in \mathcal{H}$ with $\langle\psi|\psi\rangle = 1$.

Let us try to understand this formalism using an information-theoretic point of view. Assume that we have some classical information in the form of a bit string consisting of 0's and 1's. For example, this string could have been generated by a classical coin, i.e., a binary random variable (RV) taking values 0 and 1 with probabilities p_0 and p_1, respectively. We may encode this classical information in a 2-dimensional quantum system modeled by a Hilbert space $\mathcal{H} = \mathbb{C}^2$ as follows:

$$0 \longmapsto |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad\qquad 1 \longmapsto |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{9}$$

where we denote by $\{|0\rangle, |1\rangle\}$ the standard basis of $\mathbb{C}^2$. Note that this is an orthonormal basis, i.e., $\langle 0|1\rangle = 0$, which means that that the two basis states (our quantum encodings of the classical data 0 and 1) are perfectly distinguishable.[1] However, the crucial point of Definition 2.1 is that the possible states of the quantum system can be *any normalized linear combinations* of the two distinguishable basis states $|0\rangle$ and $|1\rangle$. In physics this is known as the "superposition principle", and it is one of the fundamental features of quantum (information) theory.

A general quantum state may be written as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{10}$$

with the so-called amplitudes $\alpha, \beta \in \mathbb{C}$ satisfying $|\alpha|^2 + |\beta|^2 = 1$. This condition ensures that $\langle\psi|\psi\rangle = 1$, as required by Definition 2.1. Given a quantum state $|\psi\rangle$ as in (10), how can we retrieve the classical information encoded via (9)? This is achieved by measuring the quantum state with respect to the basis $\{|0\rangle, |1\rangle\}$.[2] Think of the measurement as asking the question "What is the probability of finding the system in the state 0, i.e., the outcome 0 of the coin flip was recorded?" The axioms of quantum theory tell us that the probabilities of obtaining the outcomes 0 and 1 are respectively given by

$$\text{Prob}(\texttt{outcome 0}) = |\langle 0|\psi\rangle|^2 = |\alpha|^2 \qquad\qquad \text{Prob}(\texttt{outcome 1}) = |\langle 1|\psi\rangle|^2 = |\beta|^2. \tag{11}$$

Note that the normalization condition $\langle\psi|\psi\rangle = 1$ ensures that $|\alpha|^2 + |\beta|^2 = 1$, i.e., $|\alpha|^2, |\beta|^2 \geq 0$ can indeed be understood as probabilities.

[1] Here, we mean distinguishable by a measurement in the sense of (11).

[2] We will only cover very rudimentary basics of the theory of measurement; there is a lot to say about measurements in quantum mechanics.

The encoding (9) of the classical information and subsequent decoding measurement (11) is *basis-dependent*. To illustrate this point, consider the following alternative encoding:

$$\begin{aligned} 0 &\longmapsto |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \\ 1 &\longmapsto |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}. \end{aligned} \tag{12}$$

This encoding uses a different orthonormal basis $\{|+\rangle, |-\rangle\}$; it is easy to check that $\langle +|-\rangle = 0$. But to retrieve the information stored in the quantum system, knowledge of the particular basis used in the encoding is crucial! To see this, assume that our quantum system is in the state $|\psi\rangle = |+\rangle$, that is, with respect to the encoding used in (12) the system is in the definite state `0`. However, if we measure $|+\rangle$ with respect to the basis $\{|0\rangle, |1\rangle\}$, we obtain the following probabilities using (11) and $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$:

$$\text{Prob}(\texttt{outcome 0}) = |\langle 0|+\rangle|^2 = \frac{1}{2} \qquad \text{Prob}(\texttt{outcome 1}) = |\langle 1|+\rangle|^2 = \frac{1}{2}. \tag{13}$$

In other words, the outcome of the measurement is completely random, and we don't learn anything about the classical information encoded via (12). This idea turns out to be quite powerful, and it can be used for cryptographic purposes [2].

Pure states represent complete knowledge of a quantum system, but the underlying assumption is that an observer has access to the whole quantum system. However, this assumption is "unphysical", and observers typically only have partial access to quantum systems. This leads to the notion of a mixed quantum state, which we now define. Recall that an operator $X \in \mathcal{L}(\mathcal{H})$ is called positive (or positive semidefinite) if $\langle\psi|X|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$. We use the shorthand $X \geq 0$ for positive operators.

Definition 2.2 (Mixed quantum states)**.** A *mixed quantum state* $\rho \in \mathcal{L}(\mathcal{H})$ is a positive semidefinite operator with unit trace: $\rho \geq 0$ and $\operatorname{tr}\rho = 1$. The set of (mixed) quantum states on $\mathcal{H}$ is denoted by

$$\mathcal{D}(\mathcal{H}) = \{\rho \in \mathcal{L}(\mathcal{H}) \colon \rho \geq 0, \operatorname{tr}\rho = 1\}. \tag{14}$$

In this more general formalism, pure states correspond to quantum states $\psi \in \mathcal{L}(\mathcal{H})$ of rank 1: they can be written (with slight overload of notation) as $\psi = |\psi\rangle\langle\psi|$ for some $|\psi\rangle \in \mathcal{H}$, and the trace condition $1 = \operatorname{tr}\psi = \operatorname{tr}|\psi\rangle\langle\psi| = \langle\psi|\psi\rangle$ ensures that $|\psi\rangle$ is normalized, as required by Definition 2.1. A mixed state $\rho \in \mathcal{D}(\mathcal{H})$ can always be written as a probabilistic mixture of pure states:

$$\rho = \sum_{i=1}^{k} p_i |\psi_i\rangle\langle\psi_i|, \tag{15}$$

where $k \geq \operatorname{rk}\rho$, the coefficients $\{p_i\}_{i=1}^k$ form a probability distribution, and $\{|\psi_i\rangle\}_{i=1}^k$ is a collection of pure states. The tuple $(\{p_i\}_{i=1}^k, \{|\psi_i\rangle\}_{i=1}^k)$ is called a quantum state ensemble realizing the mixed state ρ. Note that the vectors $|\psi_i\rangle$ need not be orthogonal to each other. For example, let $\omega = \exp(2\pi i/3)$ be a third root of unity, i.e., $\omega^3 = 1$, and consider the states

$$|\gamma_j\rangle = \frac{1}{\sqrt{2}}(|0\rangle + \omega^j|1\rangle), \quad j = 1, 2, 3. \tag{16}$$

Then we have $\frac{1}{2}\mathbb{1} = \frac{1}{3}\sum_{j=1}^{3}|\gamma_j\rangle\langle\gamma_j|$, and clearly, the $|\gamma_j\rangle$ are not orthogonal to each other. The quantum state $\frac{1}{2}\mathbb{1}$ is often called the completely mixed state. Physically, a mixed quantum state represents

uncertainty about the preparation of the quantum system. We will see in Section 2.2 below that they arise naturally from pure states when an observer only has access to part of the whole system.

A special case of (15) is the spectral decomposition of ρ, i.e.,

$$\rho = \sum_{j=1}^{\operatorname{rk}\rho} \lambda_j |e_j\rangle\langle e_j|, \tag{17}$$

where $\{\lambda_j\}_{j=1}^{\operatorname{rk}\rho}$ are the strictly positive eigenvalues of ρ, and $\{|e_j\rangle\}_{j=1}^{\operatorname{rk}\rho}$ are orthonormal eigenvectors of ρ with $\rho|e_j\rangle = \lambda_j|e_j\rangle$. Recall that the trace of an operator is equal to the sum of its eigenvalues, and eigenvalues of positive operators are non-negative. It follows that the eigenvalues $\{\lambda_j\}_j$ of a quantum state form a probability distribution. Furthermore, any classical system occupying state "i" with probability p_i can be viewed as a quantum system upon choosing some orthonormal basis $\{|i\rangle\}_i$ and considering the state $\sum_i p_i |i\rangle\langle i|$.

We conclude this section by giving the analogue of the measurement formula (11). For simplicity, consider a mixed quantum state $\rho \in \mathcal{D}(\mathbb{C}^2)$ of a qubit system, and some orthonormal basis $\{|f_0\rangle, |f_1\rangle\}$ with respect to which we measure the quantum system. Then the probabilities of obtaining the outcomes `0` and `1` associated with the basis states $|f_0\rangle$ and $|f_1\rangle$ are given by

$$\text{Prob}(\texttt{outcome 0}) = \langle f_0|\rho|f_0\rangle \qquad\qquad \text{Prob}(\texttt{outcome 1}) = \langle f_1|\rho|f_1\rangle. \tag{18}$$

Note that this formula simplifies to our earlier formula (11) if ρ is a pure state, i.e., $\rho = |\psi\rangle\langle\psi|$ for some normalized vector $|\psi\rangle \in \mathcal{H}$.

2.2 Composite systems

In quantum information theory we study the correlations of quantum systems consisting of multiple parts. We now discuss such composite quantum systems, focusing for simplicity on the bipartite case, i.e., a quantum system consisting of two parts. Let A and B be two quantum systems with associated Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. The joint state space of the quantum system AB is defined to be the tensor product of the individual Hilbert spaces:

$$\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B. \tag{19}$$

Denoting $|A| = \dim \mathcal{H}_A$ and $|B| = \dim \mathcal{H}_B$, we know from Proposition 1.4 that $\dim \mathcal{H}_{AB} = |A||B|$. Note that there is a canonical isomorphism $\mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \cong \mathcal{L}(\mathcal{H}_A) \otimes \mathcal{L}(\mathcal{H}_B)$, and mixed bipartite quantum states ρ_{AB} on AB are again positive semidefinite operators in $\mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with unit trace: $\rho_{AB} \geq 0$ and $\operatorname{tr}\rho_{AB} = 1$. Given such a bipartite state, the concept of the marginal state will be important:

Definition 2.3 (Marginal state)**.** Let $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a bipartite quantum state. Then the operators

$$\rho_A := \operatorname{tr}_B \rho_{AB} \in \mathcal{D}(\mathcal{H}_A) \qquad\qquad \rho_B := \operatorname{tr}_A \rho_{AB} \in \mathcal{D}(\mathcal{H}_B) \tag{20}$$

are called the *marginal states* of ρ_{AB} on A and B, respectively.

The marginal state on part of a quantum system models the effective state of a local observer who doesn't have access to the full system. We can use this intuition to clarify the relationship between pure states and mixed states:

Definition 2.4 (Purification of mixed states)**.** Let $\rho_A \in \mathcal{D}(\mathcal{H}_A)$ be a mixed quantum state. A *purification* of ρ_A is any pure state $|\psi\rangle_{AR} \in \mathcal{H}_A \otimes \mathcal{H}_R$ satisfying $\mathrm{tr}_R \psi_{AR} = \rho_A$.[3] Here, $\mathcal{H}_R$ is a suitable auxiliary or reference Hilbert space, often called the *purifying space*.

Purifications always exist: Let $\rho_A = \sum_{i=1}^r \lambda_i |e_i\rangle\langle e_i|_A$ be the spectral decomposition of ρ_A, and consider a basis $\{|i\rangle_R\}_{i=1}^r$ for $\mathcal{H}_R = \mathbb{C}^r$. Defining the state

$$|\psi^\rho\rangle_{AR} := \sum_{i=1}^r \sqrt{\lambda_i} |e_i\rangle_A \otimes |i\rangle_R, \tag{21}$$

we check that $|\psi^\rho\rangle_{AR}$ is a purification of ρ:

$$\begin{aligned} \mathrm{tr}_R \psi^\rho_{AR} &= \mathrm{tr}_R \Big[\textstyle\sum_{i,j} \sqrt{\lambda_i \lambda_j}\, |e_i\rangle\langle e_j|_A \otimes |i\rangle\langle j|_R \Big] && (22)\\ &= \textstyle\sum_{i,j} \sqrt{\lambda_i \lambda_j}\, |e_i\rangle\langle e_j|_A \,\mathrm{tr}\,(|i\rangle\langle j|_R) && (23)\\ &= \textstyle\sum_i \lambda_i |e_i\rangle\langle e_i|_A && (24)\\ &= \rho_A, && (25) \end{aligned}$$

where we used $\mathrm{tr}\,(|i\rangle\langle j|_R) = \langle j|i\rangle = \delta_{ij}$ in the third equality. As an example, consider the completely mixed state of a qubit,

$$\pi = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{26}$$

It is easy to see that a purification of π is given by

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle). \tag{27}$$

This state is known as EPR state (named after Einstein-Podolsky-Rosen and their discussion in the famous paper [4]), Bell state (named after John Bell, who resolved the EPR paradox [4] in the equally famous paper [1]), or simply a maximally entangled state.

Purification allows us to understand every mixed state as the marginal of a pure state on some larger system. Physically, this means that quantum states may appear mixed to us if we do not have access to the full system. We will see later in Section 2.3 that $r = \mathrm{rk}\,\rho_A$ is the minimal dimension needed for the reference Hilbert space $\mathcal{H}_R$ in order for $\mathcal{H}_A \otimes \mathcal{H}_R$ to accommodate a purification of ρ_A. Here, we record the following useful fact:

Proposition 2.5. *Let $\rho_A \in \mathcal{D}(\mathcal{H}_A)$ be a mixed state and $|\psi^\rho\rangle_{AR}$ the purification defined in (21), with $\mathcal{H}_R$ a reference system of dimension $|R| = \mathrm{rk}\,\rho_A$. Then for every other purification $|\phi\rangle_{AR'}$ of ρ_A there exists an isometry $V : \mathcal{H}_R \to \mathcal{H}_{R'}$ satisfying*

$$|\phi\rangle_{AR'} = (\mathbb{1}_A \otimes V)|\psi^\rho\rangle_{AR}. \tag{28}$$

In other words, any two isometries of a mixed state are related by an isometry acting on the purifying spaces.

Proof. See for example [20, Thm. 5.1.1]. □

[3] Remember that we slightly abuse notation by writing $\psi \equiv |\psi\rangle\langle\psi|$ for the projector corresponding to a pure state $|\psi\rangle$.

2.3 Entanglement

We now discuss the fundamental feature of quantum mechanics: entanglement. We will see that entanglement arises naturally in the question of how two quantum systems can be correlated. The typical situation we consider is that of two observers A and B, usually called Alice and Bob, who are spatially separated, by which we mean that they can only manipulate their respective quantum systems locally. Often, we also assume that they can talk to each other by sending each other classical messages (e.g., about measurement outcomes they obtained). This paradigm is referred to as *LOCC*, local operations and classical communication.

Throughout this section we restrict our attention to the bipartite case for simplicity. We first focus on the significantly simpler case of pure quantum states, assuming that we have two quantum systems A and B with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. We can divide the set of quantum states on $\mathcal{H}_A \otimes \mathcal{H}_B$ into two classes:

Definition 2.6 (Pure quantum correlations)**.** A bipartite quantum state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ is called:

- *product*, if there are states $|\phi\rangle_A \in \mathcal{H}_A$ and $|\xi\rangle_B \in \mathcal{H}_B$ such that $|\psi\rangle_{AB} = |\phi\rangle_A \otimes |\xi\rangle_B$.
- *entangled*, if it is not a product state.

Two systems A and B in a product state are completely uncorrelated. For example, any local measurement on the A system is completely independent of the state on B. On the other hand, entangled states represent correlated systems. The EPR state $|\Phi^+\rangle_{AB}$ from (27) is an example of an entangled state: it is a simple but instructive exercise to show that it cannot be written as a product state. The EPR state can be used in a remarkable protocol called teleportation, one of the most fundamental quantum communication protocols [3].

There is a handy tool to decide whether a pure state is entangled or not:

Proposition 2.7 (Schmidt decomposition)**.** *Let $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a bipartite quantum state. Then there are sets of orthonormal vectors $\{|e_i\rangle_A\}_{i=1}^r \subset \mathcal{H}_A$ and $\{|f_i\rangle_B\}_{i=1}^r \subset \mathcal{H}_B$ and (strictly) positive real numbers $\{\lambda_i\}_{i=1}^r \subset \mathbb{R}^+$ such that*

$$|\psi\rangle_{AB} = \sum_{i=1}^{r} \sqrt{\lambda_i} |e_i\rangle_A \otimes |f_i\rangle_B. \tag{29}$$

The vectors $\{|e_i\rangle_A\}$ and $\{|f_i\rangle_B\}$ are called Schmidt vectors*, the numbers $\{\lambda_i\}$ are called* Schmidt coefficients*, and the integer $r \in \mathbb{N}$ is called the* Schmidt rank *of the pure state $|\psi\rangle_{AB}$. A pure bipartite state is entangled if and only if its Schmidt rank is greater than 1. The Schmidt coefficients of a pure bipartite state are unique up to re-ordering.*

Proof. This follows directly from the singular value decomposition. For a full proof, see for example [20, Thm. 3.8.1]. □

Proposition 2.7 gives us a lot of useful information about the marginals of a pure bipartite state:

Corollary 2.8. *Let $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a bipartite quantum state with Schmidt decomposition*

$$|\psi\rangle_{AB} = \sum_{i=1}^{r} \sqrt{\lambda_i} |e_i\rangle_A \otimes |f_i\rangle_B, \tag{30}$$

as stated in Proposition 2.7. Then the marginals $\rho_A = \operatorname{tr}_B \psi_{AB} \in \mathcal{D}(\mathcal{H}_A)$ *and* $\rho_B = \operatorname{tr}_A \psi_{AB} \in \mathcal{D}(\mathcal{H}_B)$ *have spectral decompositions*

$$\rho_A = \sum_{i=1}^{r} \lambda_i |e_i\rangle\langle e_i|_A \qquad\qquad \rho_B = \sum_{i=1}^{r} \lambda_i |f_i\rangle\langle f_i|_B. \tag{31}$$

In particular, ρ_A *and* ρ_B *have the same set of non-zero eigenvalues* $\{\lambda_i\}_{i=1}^r$*, and* $r = \operatorname{rk} \rho_A = \operatorname{rk} \rho_B$*.*

Proof. This follows directly from taking partial traces in (30) and using the fact that $\{|e_i\rangle_A\}$ and $\{|f_i\rangle_B\}$ are orthonormal sets of vectors. □

Analyzing the correlations present in mixed quantum states is much more interesting (and harder) than in the pure case. We consider the following classes of correlations:

Definition 2.9 (Mixed quantum correlations)**.** A mixed quantum state $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is called:

- *product*, if there are states $\omega_A \in \mathcal{D}(\mathcal{H}_A)$ and $\sigma_B \in \mathcal{D}(\mathcal{H}_B)$ such that $\rho_{AB} = \omega_A \otimes \sigma_B$.
- *separable*, if it can be written as a convex combination of product states: there are states $\{\omega_A^{(i)}\}_{i=1}^k \subset \mathcal{D}(\mathcal{H}_A)$ and $\{\sigma_B^{(i)}\}_{i=1}^k \subset \mathcal{D}(\mathcal{H}_B)$ and a probability distribution $\{p_i\}_{i=1}^k$ such that

$$\rho_{AB} = \sum_{i=1}^{k} p_i \omega_A^{(i)} \otimes \sigma_B^{(i)}. \tag{32}$$

  Note that the set of separable states includes all product states.

- *entangled*, if it is not separable.

The class of mixed product states in Definition 2.9 is a generalization of the pure product states in Definition 2.6: If $|\psi\rangle_{AB} = |\phi\rangle_A \otimes |\chi\rangle_B$, then

$$\psi_{AB} = |\psi\rangle\langle\psi|_{AB} = |\phi\rangle\langle\phi|_A \otimes |\chi\rangle\langle\chi|_B = \phi_A \otimes \chi_B \tag{33}$$

is the corresponding mixed product state.

However, a separable state as in (32) (with $k > 1$) represents a new class of correlation, namely classical correlation stored in the classical random variable corresponding to the index i. Note that there is no quantum correlation present in such a state because of the tensor product structure of the "quantum part" of the state. The correlation in the state stems from the convex combination of product states, which is purely classical.

Given that separable states represent classical correlation, it may seem surprising that it is actually NP-hard[4] to decide whether a given quantum state is separable [9, 6]. We are therefore interested in developing tests for entanglement. An example of such a test is a so-called *necessary criterion* for entanglement. The idea is to look for an efficiently checkable property that every separably state satisfies. If a given state *fails* to have this property, we know that it must be entangled. A particularly useful example of such a criterion is based on the partial transpose that we introduced in Section 1.2.

[4]Informally speaking, a problem is in the complexity class NP (nondeterministic polynomial time) if it is easy to check a positive solution. Under the widely believed assumption that P≠NP, an NP-problem cannot be solved efficiently, that is, in polynomial time. A problem is *NP-hard* if it is at least as hard as any other problem in NP, and *NP-complete* if it is both NP-hard and a member of NP itself.

Definition 2.10 (PPT and NPT states)**.** A bipartite quantum state ρ_{AB} is called *PPT* (for positive partial transpose), if $\rho_{AB}^{T_A} \geq 0$ or equivalently $\rho_{AB}^{T_B} \geq 0$. A state ρ_{AB} is called *NPT* (non-positive partial transpose) if it is not PPT.

This definition gives rise to the PPT criterion for entanglement.

Proposition 2.11 (PPT criterion)**.** *Every separable state is PPT. An NPT state is necessarily entangled.*

Proof. To prove the first statement, let $\rho_{AB} = \sum_i p_i \omega_A^{(i)} \otimes \sigma_B^{(i)}$ be separable. Then,

$$\rho_{AB}^{T_B} = \sum_i p_i \omega_A^{(i)} \otimes \left(\sigma_B^{(i)}\right)^T, \tag{34}$$

where we used linearity of the partial transpose.

Now, transposition is a positive map (see Section 3 below), meaning that it maps positive operators to positive operators. Moreover, if $X, Y \geq 0$, then also $X \otimes Y \geq 0$, which follows directly from the spectral theorem for positive operators. Hence, $\omega_A^{(i)} \otimes (\sigma_B^{(i)})^T$ is positive for every i, since $\sigma_B^{(i)} \geq 0$ as a quantum state. It now follows from (34) that $\rho_{AB}^{T_B}$ is a convex combination of positive operators, and hence positive itself. The same argument also shows that $\rho_{AB}^{T_A} \geq 0$. The second statement follows from negating the first one. □

It turns out that in low dimensions the PPT criterion is both necessary and sufficient for separability of a state:

Proposition 2.12 ([13])**.** *Let A, B be quantum systems with $|A||B| \leq 6$. Then a bipartite state ρ_{AB} is separable if and only if it is PPT.*

Interestingly, in higher dimensions (i.e., $|A||B| > 6$) Proposition 2.12 does not hold anymore, and there are indeed entangled states that are also PPT! These are the so-called *bound entangled* states, which constitute an interesting class of quantum states sitting in between purely classical and "useful" quantum states [12].

To illustrate the PPT criterion from Proposition 2.11, let us verify that the EPR state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{35}$$

is indeed entangled; we called this state "maximally entangled" in Section 2 after all. The PPT criterion has to be evaluated on mixed states, and hence we compute $\Phi^+ = |\Phi^+\rangle\langle\Phi^+|$:

$$\Phi^+ = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{36}$$

Taking the partial transpose over B gives[5]

$$(\Phi^+)^{T_B} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}^{T_B} = \frac{1}{2}\begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}^T & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^T \\ \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}^T & \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}^T \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{37}$$

[5] It is readily checked that taking the partial transpose of Φ^+ over A leads to the same operator as in (37).

This operator has a negative eigenvalue: letting $|\psi\rangle = (0,1,-1,0)^T$, we have $(\Phi^+)^{T_B}|\psi\rangle = -\frac{1}{2}|\psi\rangle$, and hence $(\Phi^+)^{T_B} \not\geq 0$. Proposition 2.11 thus implies that Φ^+ is indeed entangled. The term "maximally entangled" stems from the fact that the marginals of Φ^+ on either system are completely mixed:

$$\mathrm{tr}_A \Phi^+ = \tfrac{1}{2}\mathbb{1} = \mathrm{tr}_B \Phi^+. \tag{38}$$

The operator $(\Phi^+)^{T_B}$ is a multiple of the so-called *swap operator* $\mathbb{F}_{AB}$, acting on product states as

$$\mathbb{F}_{AB}(|\phi\rangle_A \otimes |\chi\rangle_B) = |\chi\rangle_A \otimes |\phi\rangle_B. \tag{39}$$

Slightly rewriting the above example, we have $\mathbb{F}_{AB}(|01\rangle - |10\rangle) = -|01\rangle + |10\rangle = -(|01\rangle - |10\rangle)$.

3 Quantum channels

3.1 Introduction and physical motivation

In quantum information theory we consider maps which model processes on a given quantum system. Such maps often introduce noise into the system, which is to say the state of the system is distorted after the process, represented by the map, occurs. These maps are called *quantum operations* and we will be focused on a particular subclass of such operations, called *quantum channels*. In short, a quantum channel is a "super operator" $\Phi\colon \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ satisfying the following properties:

1. Φ is *completely positive*;
2. Φ is *trace preserving*.

Our first goal is to understand what these properties mean and why are these maps the proper way to think of quantum processes. As is typical in quantum information, we will denote the respective Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ as A and B. Thus, consider a linear map $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(B)$. Such a map induces a linear map on the matricial spaces over the domain and codomain, i.e., given such $\Phi \in \mathcal{L}(\mathcal{L}(A) : \mathcal{L}(B))$, then there is an induced map $\Phi_n \in \mathcal{L}(\mathcal{L}(\mathbb{C}^n \otimes A) : \mathcal{L}(\mathbb{C}^n \otimes B))$. The map Φ_n is called the n^{th}-*amplification* of the map Φ and is defined as

$$\sum |i\rangle\langle j| \otimes x_{ij} \mapsto \sum |i\rangle\langle j| \otimes \Phi(x_{ij}), \tag{40}$$

where $\{|i\rangle\}_{i=1}^n \subset \mathbb{C}^n$ denotes the standard basis for $\mathbb{C}^n$. The linearity of each amplification is immediate by the linearity of the initial map Φ. Given any Hilbert space $\mathcal{H}$, then there is always a set $\mathcal{L}(\mathcal{H})^+ \subset \mathcal{L}(\mathcal{H})_h$ called the *positive cone* of operators which consists of all hermitian operators ($\Phi = \Phi^\dagger$) which satisfy $\langle\psi|\Phi|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$. To say that $\mathcal{L}(\mathcal{H})^+$ is a cone simply means that $\mathcal{L}(\mathcal{H})^+ + \mathcal{L}(\mathcal{H})^+ \subset \mathcal{L}(\mathcal{H})^+$, $r\mathcal{L}(\mathcal{H})^+ \subset \mathcal{L}(\mathcal{H})^+$ for each $r \geq 0$, and $\mathcal{L}(\mathcal{H})^+ \cap -\mathcal{L}(\mathcal{H})^+ = \{0\}$. Given any $n \in \mathbb{N}$, then we may consider the algebra $\mathcal{L}(\mathbb{C}^n \otimes \mathcal{H})$ (which you may think of as $n \times n$ matrices with entries in $\mathcal{L}(\mathcal{H})$). Via this identification, each vector space $M_n(\mathcal{L}(\mathcal{H})) \simeq \mathcal{L}(\mathbb{C}^n \otimes \mathcal{H})$ has a positive cone, and we may ask whether a given map Φ also preserves these "matricial orders". This is precisely where complete positivity comes into play. A linear map $\Phi \in \mathcal{L}(\mathcal{L}(A) : \mathcal{L}(B))$ will be called *positive* if $\Phi(\mathcal{L}(A)^+) \subset \mathcal{L}(B)^+$, and will be called *completely positive* if Φ_n is positive for each $n \in \mathbb{N}$, i.e., $\Phi_n(\mathcal{L}(\mathbb{C}^n \otimes A)^+) \subset \mathcal{L}(\mathbb{C}^n \otimes B)^+$. More discussing orders of this kind will be discussed in Section 4.1.

Once again, let $\Phi \in \mathcal{L}(\mathcal{L}(A) : \mathcal{L}(B))$. Then we say Φ is *trace-preserving* if given any $\rho \in \mathcal{L}(A)$ then $\mathrm{tr}(\Phi(\rho)) = \mathrm{tr}(\rho)$. In particular, if $\rho \in \mathcal{L}(A)$ is a density, and Φ is completely positive and trace preserving, then $\Phi(\rho)$ is also a density.

Remark 3.1. A linear map $\Phi \in \mathcal{L}(M_n)$ is completely positive if and only if it is n-positive. Moreover, for linear functionals $\Phi\colon M_n \to \mathbb{C}$, a map is completely positive if and only if it is positive. Thus, for linear functionals

$$\text{complete positivity} = \text{positivity}.$$

Another instance of positivity coinciding with complete positivity is when the codomain is an "abelian C^*-algebra." One may take as an example, the $n \times n$ diagonal subalgebra of M_n.

Definition 3.2. A *quantum channel* $\Phi \in \mathcal{L}(\mathcal{L}(A) : \mathcal{L}(B))$ is defined to be a linear map which is both completely positive and trace preserving.

You may ask, why is it that quantum channels are the right thing to look at? Quantum channels are a special class of more general maps called *quantum operations*. As mentioned at the beginning of the section, quantum channels are precisely the maps that model quantum processes on quantum systems. A problem that is quite frequent is that we do not typically have the pleasure of working in a closed quantum system. This is to say, our principal system (the system we care about) will more than likely interact with an environment which influences the dynamics of the system we care about. Together, the principal system and environment form a closed quantum system and thus by the second postulate of quantum mechanics we may describe the dynamics of the total system, and this evolution is determined by a unitary on the bipartite system. To recall

Postulate II of QM: The evolution of a closed quantum system is described by a unitary transformation. Thus, the state of system at time t_o, $|\xi_o\rangle$, is related to the state at time t_1, $|\xi_1\rangle$, by a unitary operator U;

$$|\xi_1\rangle = U|\xi_o\rangle. \tag{41}$$

Equivalently, we may describe the dynamics of a closed system by the Hamiltonian:
Postulate II′ of QM: The time evolution of the state of a closed quantum system is described by the Schrödinger equation

$$i\hbar\frac{\mathrm{d}}{\mathrm{dt}}|\xi\rangle = H|\xi\rangle. \tag{42}$$

Thus, if we were to let P and E denote our principal system and environment, respectively, and if our bipartite system is prepared in the product state $\rho_P \otimes \rho_E$, then we may consider the state of the principal system after applying a unitary $U \in \mathcal{L}(\mathcal{L}(P \otimes E) : \mathcal{L}(P \otimes E))$, followed by tracing out the environment:

$$\mathrm{tr}_E(U(\rho_P \otimes \rho_E)U^\dagger) =: \Phi(\rho). \tag{43}$$

The trouble is that the initial state ρ of the principal system, and the final state $\Phi(\rho)$ of the principal system, may not be related by a unitary. This is precisely where the *operator-sum formulation* comes into play. Given any quantum operation, we consider the operator-sum representation (think of Kraus' theorem below) which can describe nearly closed systems which are weakly coupled to the environment, but also open systems which are strongly coupled to their environment. We wish to elaborate on this point a bit more.
Fix an orthonormal basis $\{|i_E\rangle\} \subset E$ for the environment, which for simplicity, we will assume is in the initial state $|0_E\rangle\langle 0_E|$. We lose no information by assuming E begins in a pure state, since by the

discussion following Definition 2.4, we may always assume if the initial state of E is mixed, then it is necessarily the reduced state of a pure state on a larger system. Thus, Equation (43) becomes

$$\begin{aligned}\Phi(\rho) &:= \mathrm{tr}_E(U(\rho_P \otimes |0_E\rangle\langle 0_E|)U^\dagger) && (44)\\ &= \sum (\mathbb{1}_P \otimes \langle i_E|)\left[U(\rho_P \otimes |0_E\rangle\langle 0_E|)U^\dagger\right](\mathbb{1}_P \otimes |i_E\rangle) && (45)\\ &= \sum (\mathbb{1}_P \otimes \langle i_E|)\left[U(\mathbb{1}_p \otimes |0_E\rangle)(\rho_P \otimes \mathbb{1}_E)(\mathbb{1}_P \otimes \langle 0_E|)U^\dagger\right](\mathbb{1}_P \otimes |i_E\rangle) && (46)\\ &= \sum A_i \rho A_i^\dagger, && (47)\end{aligned}$$

where $A_i := (\mathbb{1}_p \otimes \langle i_E|)U(\mathbb{1}_P \otimes |i_E\rangle) \in \mathcal{L}(\mathcal{L}(P))$ for each i. Equation (46) is known as the *operator-sum representation* or *Kraus representation* of Φ. The operators $\{A_i\} \subset \mathcal{L}(\mathcal{L}(P))$ are known as the *operator elements* or *Kraus operators* of Φ.

A *quantum operation* is defined to be a map $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(B)$ which is completely positive, given any density ρ, $0 \le \mathrm{tr}(\Phi(\rho)) \le 1$, and is convex on the cone of density operators. This is to say, if $\{\rho_i\}$ a subset of densities with $\{p_i\}$ a probability distribution, then

$$\Phi\left(\sum_i p_i \rho_i\right) = \sum p_i \Phi(\rho_i).$$

The assumption of complete positivity is motivated by the following: consider a quantum operation $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(A)$ and suppose R is another quantum system. If we consider the Hilbertian tensor product $R \otimes A$, then one would expect that the simple tensor $\mathrm{id}_R \otimes \Phi\colon \mathcal{L}(R \otimes A) \to \mathcal{L}(R \otimes A)$ is a quantum operation, and thus models a process on the system $R \otimes A$. In particular, the map $\mathrm{id}_R \otimes \Phi$ should be positive. If $\dim R$ is at least as large as $\dim A$, then this is equivalent to complete positivity of Φ. In saying that the maps we consider should model quantum processes, it should happen that the trace of a quantum operation should yield the probability that said process occurs. In particular, if our system is in the initial state ρ, then

$$0 \le \mathrm{tr}(\Phi(\rho)) \le 1, \tag{48}$$

should give the probability that the process representing Φ occurs.

3.2 Some fundamental theorems

Our goal in this subsection will be to discuss three fundamental theorems relative to the study of quantum channels. These are *Choi's isomorphism theorem, Choi-Krauss representation,* and *Stinespring's theorem*.

We begin with the *Choi-Jamiolkowski isomorphism*, which we will refer to as the "Choi-isomorphism." We first make the following comments: suppose we start with a super operator $\Phi \in \mathcal{L}(\mathcal{L}(A) : \mathcal{L}(B))$ and fix an o.n.b. $\{|i\rangle\}_{i=1}^{\dim A}$. We will be concerned with a matrix representation of Φ which is able to "detect" many properties of Φ. Fixing an orthonormal basis $\{|i\rangle\}_{i=1}^{\dim A}$, we define the *Choi matrix* of the super operator Φ as the operator $\mathrm{Ch}(\Phi) := \sum_{ij} |i\rangle\langle j| \otimes \Phi(|i\rangle\langle j|)$. Consider the scaled maximally entangled state $|\gamma\rangle := \sum_i |ii\rangle \in A \otimes A$, and let $\gamma := |\gamma\rangle\langle\gamma|$. We then write the Choi matrix of Φ as

$$\tau := (\mathrm{id}_A \otimes \Phi)(\gamma). \tag{49}$$

Theorem 3.3. *There is a linear isomorphism between the vector spaces*

$$\begin{aligned}\Psi\colon \mathcal{L}(\mathcal{L}(A):\mathcal{L}(B)) &\to \mathcal{L}(A\otimes B)\\ \Phi &\mapsto \tau = (\mathbb{1}_A\otimes\Phi)(\gamma),\end{aligned}$$

with inverse given by

$$\begin{aligned}\Psi^{-1}\colon \mathcal{L}(A\otimes B) &\to \mathcal{L}(\mathcal{L}(A):\mathcal{L}(B))\\ \tau &\mapsto \left(\Phi\colon \rho\mapsto \mathrm{tr}_A((\rho^T\otimes\mathbb{1}_B)(\tau))\right),\end{aligned}$$

where $\Phi\in\mathcal{L}(\mathcal{L}(A):\mathcal{L}(B))$, and $\{|i\rangle\}_{i=1}^{dim\,A}$ is a fixed orthonormal basis for the Hilbert space A.

Proof. We will show that Ψ has both a left inverse and a right inverse. To begin, let $\Phi\in\mathcal{L}(\mathcal{L}(A):\mathcal{L}(B)), \rho\in\mathcal{L}(A)$, and consider the output $\Psi(\Phi)=\tau$. It then follows

$$\begin{aligned}\mathrm{tr}_A((\rho^T\otimes\mathbb{1}_B)(\tau)) &= \mathrm{tr}_A(\sum_{ij}(\rho^T|i\rangle\langle j|\otimes\Phi(|i\rangle\langle j|)))\\ &= \sum_{ij}\langle j|\rho^T|i\rangle\Phi(|i\rangle\langle j|)\\ &= \sum_{ij}\langle i|\rho|j\rangle\Phi(|i\rangle\langle j|)\\ &= \sum_{ij}\Phi(\rho_{ij}|i\rangle\langle j|)\\ &= \Phi\left(\sum_{ij}\rho_{ij}|i\rangle\langle j|\right)\\ &= \Phi(\rho).\end{aligned}$$

Thus, Ψ is left-invertible and therefore injective. Conversely, let $\tau\in\mathcal{L}(A\otimes B)$. We want to construct a super operator $\Phi\in\mathcal{L}(\mathcal{L}(A):\mathcal{L}(B))$ such that $\Phi\mapsto_\Psi\tau$. To this end, let $\{|j\rangle_B\}_{j=1}^{\dim B}$ be an arbitrary (orthonormal) basis for B, and suppose given a vector $|\varphi\rangle\in A\otimes B$, we write it as

$$|\varphi\rangle = \sum_{ij}\rho_{ij}|i\rangle\otimes|j\rangle_B.$$

Define the map $V: A\to B$ as $V\coloneqq\sum_{ij}\rho_{ij}|j\rangle_B\langle i|$. If we then consider $|\gamma\rangle\in A\otimes A$, as above, we have

$$\begin{aligned}(\mathbb{1}_A\otimes V)(|\gamma\rangle) &= (\mathbb{1}_A\otimes V)(\sum_k|kk\rangle) &(50)\\ &= \sum_{ijk}|k\rangle\rho_{ij}\langle i|k\rangle|j\rangle_B &(51)\\ &= \sum_{ij}\rho_{ij}|i\rangle|j\rangle_B &(52)\\ &= |\varphi\rangle. &(53)\end{aligned}$$

Thus, we may realize the vector $|\varphi\rangle=(\mathbb{1}_A\otimes V)(|\gamma\rangle)\in A\otimes B$.

With $\tau \in \mathcal{L}(A \otimes B)$, we may write $\tau = \sum_i |\xi_i\rangle\langle\eta_i|$ where both $|\xi_i\rangle, |\eta_i\rangle$ are pure states, such that $|\xi_i\rangle \neq |\eta_i\rangle$. Using our remark above, we know for each i there exists operators $V_i, W_i : A \to B$ such that

$$|\xi_i\rangle = (\mathbb{1}_A \otimes V_i)(|\gamma\rangle)$$
$$|\eta_i\rangle = (\mathbb{1}_A \otimes W_i)(|\gamma\rangle).$$

Rewriting τ we obtain $\tau = \sum_i |\xi_i\rangle\langle\eta_i| = \sum_i (\mathbb{1}_A \otimes V_i)|\gamma\rangle\langle\gamma|(\mathbb{1}_A \otimes W_i)^\dagger$. Define the super operator $\Phi(\cdot) := \sum_i V_i(\cdot)W_i^\dagger : \mathcal{L}(A) \to \mathcal{L}(B)$, and let us compute $\Psi(\Phi)$:

$$\begin{aligned}
\Psi(\Phi) &= (\mathrm{id}_A \otimes \Phi)(\gamma) \\
&= \sum_{ij} |i\rangle\langle j| \otimes \Phi(|i\rangle\langle j|) \\
&= \sum_{ij} |i\rangle\langle j| \otimes \sum_\ell V_\ell |i\rangle\langle j| W_\ell^\dagger \\
&= \sum_{ij\ell} (\mathbb{1}_A \otimes V_\ell)|ii\rangle\langle jj|(\mathbb{1}_A \otimes W_\ell)^\dagger \\
&= \sum_\ell (\mathbb{1}_A \otimes V_\ell)|\gamma\rangle\langle\gamma|(\mathbb{1}_A \otimes W_\ell)^\dagger \\
&= \tau
\end{aligned}$$

Thus, Ψ is surjective, and with the first part of our proof we have Ψ is a linear isomorphism. □

One reason the Choi isomorphism is so powerful is that we are able detect various properties of Φ by looking at τ! Some properties appear as the following proposition:

Proposition 3.4. *Let $\Phi \in \mathcal{L}(\mathcal{L}(A) : \mathcal{L}(B))$ be a super operator, and let $\tau \in \mathcal{L}(A \otimes B)$ be the induced Choi matrix. Then the following statements hold:*

1. *$\Phi(\rho)^\dagger = \Phi(\rho^\dagger)$ if and only if $\tau^\dagger = \tau$.*
2. *Φ is completely positive if and only if τ is positive semidefinite.*
3. *Φ is trace preserving if and only if $\mathrm{tr}_B(\tau) = \mathbb{1}_A$.*
4. *Φ is unital if and only if $\mathrm{tr}_A(\tau) = \mathbb{1}_B$.*

Theorem 3.5 (Kraus Representation)**.** *A linear map $\Phi : \mathcal{L}(A) \to \mathcal{L}(B)$ is completely positive if and only if there exist operators $\{V_i\}_{i=1}^r, V_i : A \to B$ such that*

$$\Phi(\cdot) = \sum_i V_i(\cdot)V_i^\dagger. \tag{54}$$

Proof. Given a family of operators $\{V_i\}_{i=1}^r, V_i : A \to B$, then for each $i \in [r] := \{1, \ldots, r\}$, the action $V_i(\cdot)V_i^\dagger$ is completely positive (it is an automorphism). In particular, given any $\rho \in \mathcal{L}(A)^+$ then $V_i \rho V_i^\dagger \in \mathcal{L}(B)^+$. Since the set of positive elements forms a cone (See Section 3.1) it follows $\sum_{i=1}^r V_i \rho V_i^\dagger \in \mathcal{L}(B)^+$. Thus, if $\Phi(\cdot) := \sum V_i(\cdot)V_i^\dagger$, then Φ is completely positive.

Conversely, suppose $\Phi : \mathcal{L}(A) \to \mathcal{L}(B)$ is completely positive. Then by our proposition above we know the Choi matrix $\tau \in \mathcal{L}(A \otimes B)$ is positive semidefinite. By the Spectral Theorem we may write

$\tau = \sum_{i=1}^{r} \lambda_i |\xi_i\rangle\langle\xi_i|$, which by absorbing the λ_i's into the vectors, we will write as $\sum_i |\tilde{\xi}_i\rangle\langle\tilde{\xi}_i|$. Via our methods in Theorem 3.3 we know for each $i \in [r]$ there exists an operator $V_i : A \to B$ such that $(\mathbb{1}_A \otimes V_i)|\gamma\rangle = |\tilde{\xi}_i\rangle$. Let $\rho \in \mathcal{L}(A)$. Then it must follow $\Phi(\rho) = \mathrm{tr}_A((\rho^T \otimes \mathbb{1}_B)(\tau)) = \Psi^{-1}(\tau)$. Let us compute and see what happens:

$$\begin{aligned}
\mathrm{tr}_A((\rho^T \otimes \mathbb{1}_B)(\tau)) &= \mathrm{tr}_A((\rho^T \otimes \mathbb{1}_B)(\sum (\mathbb{1}_A \otimes V_i)|\gamma\rangle\langle\gamma|(\mathbb{1}_A \otimes V_i)^\dagger)) \\
&= \sum_{ij} \mathrm{tr}_A((\rho^T |j\rangle \otimes V_i|j\rangle)(\langle j| \otimes \langle j|V_i^\dagger)) \\
&= \sum_{ijk} \mathrm{tr}_A(\rho^T |j\rangle\langle k| \otimes V_i|j\rangle\langle k|V_i^\dagger) \\
&= \sum_{ijk} \langle j|\rho|k\rangle V_i|j\rangle\langle k|V_i^\dagger \\
&= \sum_i V_i(\sum_{jk} \rho_{jk}|j\rangle\langle k|)V_i^\dagger \\
&= \sum_i V_i \rho V_i^\dagger.
\end{aligned}$$

This proves $\Phi(\rho) = \sum_{i=1}^{r} V_i \rho V_i^\dagger$, which finishes the proof. □

Theorem 3.6 (Stinespring's Dilation Theorem (finite-dimensional))**.** *Let $\Phi : \mathcal{L}(A) \to \mathcal{L}(B)$ be a super operator. Then if Φ is completely positive, there exists a Hilbert space C, and a linear operator $V : A \to B \otimes C$ such that*

$$\Phi(\cdot) = \mathrm{tr}_C V(\cdot)V^\dagger. \tag{55}$$

Proof. Suppose Φ is completely positive, which by Theorem 3.5, we may write as $\Phi(\cdot) = \sum_{i=1}^{r} V_i(\cdot)V_i^\dagger$, for some operators $V_i : A \to B, i \in [r]$. Consider the Hilbert space $\mathbb{C}^r$, which we denote by C, and let $\{|i_C\rangle\}_i \subset C$ denote the standard basis. Define the operator $V : A \to B \otimes C$ as $V|\xi\rangle := \sum_i V_i|\xi\rangle \otimes |i_C\rangle$. Consider $\xi := |\xi\rangle\langle\xi| \in \mathcal{L}(A)$. We verify $\Phi(\xi) = \mathrm{tr}_C(V\xi V^\dagger)$, which will prove the result after extending linearly. We see

$$\begin{aligned}
\mathrm{tr}_C(V\xi V^\dagger) &= \mathrm{tr}_C((\sum_i V_i|\xi\rangle \otimes |i_C\rangle)(\sum \langle\xi|V_i^\dagger \otimes \langle i|_C)) \\
&= \mathrm{tr}_C(\sum_{ij} V_i \xi V_j^\dagger \otimes |i_C\rangle\langle j_C|) \\
&= \sum_{ij} V_i \xi V_j^\dagger \langle j_C|i_C\rangle \\
&= \sum_i V_i \xi V_i^\dagger \\
&= \Phi(\rho).
\end{aligned}$$

This finishes the proof after extending linearly. □

3.3 Special classes of quantum channels

In this section we discuss a few special classes of quantum channels. Recall that a quantum channel $\Phi : \mathcal{L}(A) \to \mathcal{H}(B)$ maps the operators acting on a Hilbert space A to those acting on a Hilbert space B.

3.3.1 Entanglement-breaking channels

Recall from Section 3.2 that the Choi matrix τ_{AB}^{Φ} of a channel $\Phi\colon \mathcal{L}(A) \to \mathcal{H}(B)$ is defined as

$$\tau_{AB}^{\Phi} := (\mathrm{id}_A \otimes \Phi)(\gamma_{AA'}), \tag{56}$$

where $|\gamma\rangle_{AA'} = \sum_{i=1}^{|A|} |i\rangle_A |i\rangle_{A'}$ is the (unnormalized) maximally entangled state between A and $A' \cong A$, defined in terms of some o.n.b. $\{|i\rangle\}_{i=1}^{|A|}$ of $\mathcal{H}_A$. A quantum channel models noise affecting our quantum system, and hence a quantum channel acting on part of an entangled bipartite quantum system will generally reduce the correlation, i.e., the channel output will be less entangled than the input. In this operational picture, we can understand the Choi matrix τ_{AB}^{Φ} of (56) as the channel Φ acting on half of a maximally entangled state $|\gamma\rangle$. The noiser the channel Φ, the less entanglement is left in τ^{Φ}.[6] In the extreme case, a quantum channel may destroy any initial entanglement, i.e., the output is a separable and void of any quantum correlation between the systems A and B. Such channels are known as *entanglement-breaking* and formally defined as follows:

Definition 3.7 (Entanglement-breaking channels)**.** A quantum channel $\Phi\colon \mathcal{L}(A) \to \mathcal{H}(B)$ is called *entanglement-breaking* (EB) if $(\mathrm{id}_R \otimes \Phi)(\rho_{RA})$ is separable for any bipartite quantum state ρ_{RA}, where R is an auxiliary quantum system of arbitrary size.

The following result provides some useful criteria to check whether a channel is EB.

Proposition 3.8 (Equivalent conditions for EB channels)**.** *For a quantum channel $\Phi\colon \mathcal{L}(A) \to \mathcal{H}(B)$, the following conditions are equivalent:*

(i) *Φ is EB.*

(ii) *τ_{AB}^{Φ} is separable.*

(iii) *Φ has a Kraus representation consisting of rank-1 operators.*

(iv) *There exist operators $\{E_A^i\}$ on $\mathcal{H}_A$ satisfying $E_A^i \geq 0$ for all i and $\sum_i E_A^i = \mathbb{1}_A$, as well as quantum states $\{\sigma_B^i\}$ on $\mathcal{H}_B$, such that the channel Φ acts on an input state ρ_A as*

$$\Phi(\rho_A) = \sum\nolimits_i \mathrm{tr}\left(E_A^i \rho_A\right) \sigma_B^i. \tag{57}$$

Proof. (i) $\Rightarrow$ (ii): Follows directly from the definition.
(ii) $\Rightarrow$ (iii): By assumption, τ_{AB}^{Φ} is separable, so there are pure quantum states $\{|\psi^i\rangle_A\}$ and $\{|\phi^i\rangle_B\}$ and a probability distribution $\{p_i\}$ such that we can write the normalized Choi state as

$$\frac{1}{|A|}\tau_{AB}^{\Phi} = \sum_i p_i \psi_A^i \otimes \phi_B^i. \tag{58}$$

We define the rank-1 operators

$$K_i := \sqrt{|A| p_i}\, |\phi^i\rangle_B \overline{\langle\psi^i|}_A, \tag{59}$$

[6] This idea can be used to define a quantity called the *entanglement fidelity* of a quantum channel, which measures how far the channel Φ is from the identity channel; see [20, Sec. 9.5.2] for details.

where $\overline{\langle\psi^i|}_A$ is the entry-wise complex conjugate of $\langle\psi^i|_A$ with respect to the same basis for A used to define the state $|\gamma\rangle_{AA'}$ in the Choi matrix τ^Φ. We claim that $\{K_i\}$ is the desired rank-1 Kraus decomposition of Φ.

First, we verify that $\{K_i\}$ is indeed a Kraus representation of *some* channel:

$$\sum_i K_i^\dagger K_i = |A| \sum_i p_i \overline{|\psi^i\rangle}_A \langle\phi^i|\phi^i\rangle_B \overline{\langle\psi^i|}_A = |A| \sum_i p_i \overline{\psi_A^i} = |A| \overline{\sum_i p_i \psi_A^i} \overset{(*)}{=} |A| \overline{\frac{1}{|A|} \mathbb{1}_A} = \mathbb{1}_A. \tag{60}$$

The equality $(*)$ follows from taking the partial trace over B in (58) and using $\mathrm{tr}_B\, \tau_{AB}^\Phi = \mathbb{1}_A$ from Proposition 3.4. Second, we compute the Choi matrix of the channel defined via $\{K_i\}$:

$$\begin{aligned}
\sum_i (\mathbb{1}_A \otimes K_i)|\gamma\rangle\langle\gamma|_{AA'} (\mathbb{1}_A \otimes K_i)^\dagger &= \sum_{i,j,k} |j\rangle\langle k|_A \otimes K_i |j\rangle\langle k|_{A'} K_i^\dagger && (61)\\
&= |A| \sum_{i,j,k} p_i |j\rangle\langle k|_A \otimes \overline{\langle\psi^i|j\rangle} \langle k|\overline{\psi^i}\rangle \phi_B^i && (62)\\
&= |A| \sum_{i,j,k} p_i \, \langle j|\overline{\psi_A^i}|k\rangle \, |j\rangle\langle k|_A \otimes \phi_B^i && (63)\\
&= |A| \sum_i p_i \underbrace{\sum_{j,k} \langle j|\overline{\psi_A^i}|k\rangle \, |j\rangle\langle k|_A}_{= \left(\overline{\psi}_A^i\right)^T = (\psi_A^i)^\dagger = \psi_A^i} \otimes \phi_B^i && (64)\\
&= |A| \sum_i p_i \psi_A^i \otimes \phi_B^i && (65)\\
&= \tau_{AB}^\Phi, && (66)
\end{aligned}$$

and hence $\{K_i\}$ is indeed a Kraus representation of Φ by Theorem 3.3.

(iii) $\Rightarrow$ (iv): We can write the rank-1 operators K_i as $K_i = |\chi^i\rangle_B \langle\omega^i|_A$ for some normalized vectors $|\chi^i\rangle_B$ (i.e., $\langle\chi|\chi\rangle_B = 1$) and arbitrary vectors $|\omega^i\rangle_A$. Note that the normalization conditions for $\{K_i\}$ and the vectors $\{|\chi^i\rangle\}$ imply that

$$\mathbb{1}_A = \sum_i K_i^\dagger K_i = \sum_i |\omega^i\rangle_A \langle\chi^i|\chi^i\rangle_B \langle\omega^i|_A = \sum_i |\omega^i\rangle\langle\omega^i|_A, \tag{67}$$

and so the positive semidefinite operators $\omega_A^i = |\omega^i\rangle\langle\omega^i|_A \geq 0$ satisfy the desired normalization condition $\sum_i \omega_A^i = \mathbb{1}_A$. With these choices,

$$\begin{aligned}
\Phi(\rho_A) &= \sum_i K_i \rho_A K_i^\dagger && (68)\\
&= \sum_i \langle\omega^i|\rho_A|\omega^i\rangle \, |\chi^i\rangle\langle\chi^i|_B && (69)\\
&= \sum_i \mathrm{tr}(\omega_A^i \rho_A) \, |\chi^i\rangle\langle\chi^i|_B, && (70)
\end{aligned}$$

which is the advertised action of Φ on an input state ρ_A.

(iv) ⇒ (i): Let $\{E_A^i\}$ be a collection of positive semidefinite operators satisfying $\sum_i E_A^i = \mathbb{1}_A$, and let $\{\sigma_B^i\}$ be states such that $\Phi(\rho_A) = \sum_i \operatorname{tr}(E_A^i \rho_A)\sigma_B^i$. For a bipartite state ρ_{RA} with R arbitrary, we have

$$(\mathrm{id}_R \otimes \Phi)(\rho_{RA}) = \sum_i \operatorname{tr}_A\left[(\mathbb{1}_R \otimes E_A^i)\rho_{RA}\right] \otimes \sigma_B^i \tag{71}$$

$$= \sum_i \operatorname{tr}_A\left[\left(\mathbb{1}_A \otimes \sqrt{E_A^i}\right) \rho_{RA} \left(\mathbb{1}_A \otimes \sqrt{E_A^i}\right)\right] \otimes \sigma_B^i. \tag{72}$$

Here we used the fact that E_A^i are positive operators with a unique positive square root $\sqrt{E_A^i}$. The second equality then follows using (partial) cyclicity of the partial trace. Defining the quantum states

$$\omega_A^i := \frac{1}{p_i} \operatorname{tr}_A\left[\left(\mathbb{1}_A \otimes \sqrt{E_A^i}\right) \rho_{RA} \left(\mathbb{1}_A \otimes \sqrt{E_A^i}\right)\right] \geq 0 \tag{73}$$

with $p_i := \operatorname{tr}(E_A^i \rho_A)$ such that $\operatorname{tr}\omega_A^i = 1$, it follows that $(\mathrm{id}_R \otimes \Phi)(\rho_{RA}) = \sum_i p_i \omega_A^i \otimes \sigma_B^i$ is separable for any ρ_{RA}, which concludes the proof. □

The action of an EB channel Φ as in Proposition 3.8 is called the *Holevo form* of an EB channel, and it can be viewed in operational terms as a "measure-and-prepare" procedure. To understand this terminology, we need to take a quick detour and revisit the idea of measuring a quantum system. At the end of Section 2.1 we introduced a specific measurement of a quantum system A defined in terms of an o.n.b. $\{|e_i\rangle_A\}_{i=1}^{|A|}$: Given a mixed quantum state $\rho_A \in \mathcal{L}(A)$, the outcome $\mathtt{i}$ associated to the basis vector $|e_i\rangle_A$ is obtained with probability

$$\Pr(\mathtt{i}) = \langle e_i|\rho_A|e_i\rangle = \operatorname{tr}(|e_i\rangle\langle e_i|_A \rho_A). \tag{74}$$

Note that the projectors $|e_i\rangle\langle e_i|_A$ satisfy the following two properties: $|e_i\rangle\langle e_i|_A \geq 0$ for all i and $\sum_i |e_i\rangle\langle e_i|_A = \mathbb{1}_A$, which follows from the fact that $\{|e_i\rangle_A\}_{i=1}^{|A|}$ is an o.n.b. We can generalize this in the following way:

Definition 3.9 (POVMs)**.** A *positive operator-value measure* (POVM) on a quantum system A is a collection of positive operators or *effect operators* $E_A^i \geq 0$ for $i = 1, \ldots, k$ such that

$$\sum_{i=1}^k E_A^i = \mathbb{1}_A.$$

A POVM $\{E_A^i\}_{i=1}^k$ corresponds to a quantum measurement in the following way: the possible outcomes $\{1, \ldots, \mathtt{k}\}$ of the measurement are associated with the effect operators $\{E_1, \ldots, E_k\}$, and given a state ρ_A of the quantum system A the measurement $\{E_A^i\}_{i=1}^k$ yields outcome $\mathtt{i}$ with probability

$$\Pr(\mathtt{i}) = \operatorname{tr}(E_A^i \rho_A). \tag{75}$$

Note that both E_A^i and ρ_A are positive semidefinite, and hence $\operatorname{tr}(E_A^i \rho_A)$ is always non-negative.[7] Furthermore, since $\sum_i E_A^i = \mathbb{1}_A$, we have by linearity of the trace that

$$\sum\nolimits_i \Pr(\mathtt{i}) = \sum\nolimits_i \operatorname{tr}(E_A^i \rho_A) = \operatorname{tr}\left[\left(\sum\nolimits_i E_A^i\right)\rho_A\right] = \operatorname{tr}\rho_A = 1. \tag{76}$$

[7] To see this, let $X, Y \geq 0$. Then also $\sqrt{X}\,Y\sqrt{X} \geq 0$, and $\operatorname{tr}(XY) = \operatorname{tr}(\sqrt{X}\,Y\sqrt{X}) \geq 0$ as the trace of a positive operator.

Hence, the defining properties of a POVM ensure that the numbers $\Pr(\mathtt{i}) = \operatorname{tr}(E_A^i \rho_A)$ indeed form a probability distribution.

Let now $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(B)$ be an EB quantum channel. Then by Proposition 3.8(iv) there exist a POVM $\{E_A^i\}$ on A and quantum states $\{\sigma_B^i\}$ on B such that

$$\Phi(\rho_A) = \sum_i \operatorname{tr}(E_A^i \rho_A)\sigma_B^i. \tag{77}$$

We can now understand the action of the channel as follows:

(a) Measure the system A with respect to the POVM $\{E_A^i\}$;

(b) Upon obtaining the outcome $\mathtt{i}$, prepare the state σ_B^i.

This procedure explains the name "measure-and-prepare" channel. Since measurements destroy entanglement, such channels *must* be entanglement-breaking. The pleasing result of Proposition 3.8 is that *every* entanglement-breaking channel can be understood in this way.

3.3.2 PPT channels

Recall from the discussion in Section 2.3 that checking whether a mixed state is separable is NP-hard. The PPT criterion from Definition 2.10 and Proposition 2.11 is a useful tool to detect entangled states. Since a channel is entanglement-breaking iff its Choi state is separable by Proposition 3.8(ii), we can also apply this criterion to detect whether a channel is (not) entanglement-breaking. Even though channels whose Choi matrix is PPT are not necessarily entanglement-breaking (see the discussion after Proposition 2.12), they are nevertheless interesting in their own right [14, 11, 10], and hence we make the following definition which should we compared with Definition 3.7:

Definition 3.10 (PPT channels)**.** A quantum channel $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(B)$ is called PPT if $(\mathrm{id}_R \otimes \Phi)(\rho_{RA})$ is PPT for any bipartite quantum state ρ_{RA}, where R is an auxiliary quantum system of arbitrary size.

Similar to our discussion of entanglement-breaking channels, we now prove the following set of equivalent conditions for PPT channels.

Proposition 3.11 (Equivalent conditions for PPT channels)**.** *For a quantum channel $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(B)$, the following are equivalent:*

(i) *Φ is PPT.*

(ii) *τ_{AB}^{Φ} is PPT.*

(iii) *$T \circ \Phi$ is CP, where T denotes the transposition map.*

Proof. (i) $\Rightarrow$ (ii): True by definition.
(ii) $\Rightarrow$ (iii): We have

$$\begin{aligned}(\mathrm{id}_A \otimes T \circ \Phi)(\gamma) &= (\mathrm{id}_A \otimes T)(\mathrm{id}_A \otimes \Phi)(\gamma) &(78)\\ &= (\mathrm{id}_A \otimes T)(\tau_{AB}^{\Phi}) &(79)\\ &= (\tau_{AB}^{\Phi})^{T_B} \geq 0, &(80)\end{aligned}$$

by assumption of τ_{AB}^{Φ} being PPT. Since $(\mathrm{id}_A \otimes T \circ \Phi)(\gamma)$ is the positive Choi state of the map $T \circ \Phi$, the latter is CP by Proposition 3.4.
(iii) $\Rightarrow$ (i): Let ρ_{RA} be arbitrary. Since $T \circ \Phi$ is CP by assumption, the map $\mathrm{id}_R \otimes T \circ \Phi$ is positive for arbitrary R. It follows that

$$0 \leq (\mathrm{id}_R \otimes T \circ \Phi)(\rho_{RA}) = (\mathrm{id}_R \otimes T)(\mathrm{id}_R \otimes \Phi)(\rho_{RA}) = [(\mathrm{id}_R \otimes \Phi)(\rho_{RA})]^{T_B}, \tag{81}$$

which means that Φ is PPT. □

3.3.3 Covariant channels

We now want to study symmetries of quantum channels. For simplicity, we restrict to quantum channels mapping between identical input and output systems. In quantum mechanics, symmetries are typically assumed to be unitary operators. Denoting the set of unitary operators on A by $\mathcal{U}(A)$, we call a unitary $U_A \in \mathcal{U}(A)$ a *symmetry of a quantum channel* $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(A)$ if

$$\Phi(U_A \cdot U_A^\dagger) = U_A \Phi(\cdot) U_A^\dagger. \tag{82}$$

We will particularly be interested in the case where the symmetries of a quantum channel form a *group*, which for our purposes can simply be understood as a subset $G \subset \mathcal{U}(A)$ of the unitaries on A with the following special structure:

(G1) If $U, V \in G$, then also $UV \in G$ (*composing two symmetries yields another symmetry*).

(G2) $\mathbb{1}_A \in G$ (*the identity $\mathbb{1}_A$ is a symmetry*).

(G3) $U^{-1} \in G$ for all $U \in G$ (*every element in G has an inverse*).

In (G3), U^{-1} denotes the inverse of U as an operator (or matrix, since we only work in finite dimensions), and we have $UU^{-1} = U^{-1}U = \mathbb{1}_A$. Note that the following additional property is automatically satisfied, since matrix multiplication (or equivalently, operator composition) is associative:

(G4) $(UV)W = U(VW)$ for $U, V, W \in G$ (*associativity of composition*).

More abstractly, an arbitrary set G with a 'multiplication' operation $G \times G \to G$ is called a *group* if it satisfies the formal analogues of the axioms (G1)-(G4).[8] The mathematical subfield studying the rich theory of these algebraic structures is called group theory [15]. In this more general framework of abstract groups, one may view a group G as a subgroup of the unitaries $\mathcal{U}(A)$ by considering a map $\varphi\colon G \to \mathcal{U}(A)$ satisfying $\varphi(gh) = \varphi(g)\varphi(h)$ for all $g, h \in G$. Intuitively, φ realizes the elements of an abstract group G as unitary operators acting on a Hilbert space in a way that preserves the structure of G. This is a special case of a *representation* of an abstract group on a vector space, and leads to the equally rich and important field of representation theory [5].

Armed with this terminology, we now define the following class of quantum channels.

[8] A group $(G, \circ)$ is a non-empty set G together with a binary operation $\circ\colon G \times G \to G$ satisfying the following properties:

- Associativity: $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$ for all $g_1, g_2, g_3 \in G$.
- Neutral element: there exists $e \in G$ such that $e \circ g = g$ for all $g \in G$.
- Inverses: for all $g \in G$ there exists $h \in G$ with $g \circ h = e$.

Definition 3.12 (Covariant channels)**.** Let $G \subset \mathcal{U}(A)$ be a group, i.e., a subset of the unitaries on A satisfying conditions (G1), (G2), and (G3) above. A quantum channel $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(A)$ is called *covariant with respect to* G if

$$\Phi(U_A \cdot U_A^\dagger) = U_A \Phi(\cdot) U_A^\dagger \qquad \text{for all } U_A \in G. \tag{83}$$

Let us look at some examples of covariant channels. Recall the Pauli matrices

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \equiv I \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \equiv X \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \equiv Y \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \equiv Z, \tag{84}$$

which satisfy the following properties:

$$\sigma_j = \sigma_j^\dagger \in M_2(\mathbb{C}), \quad \sigma_j^2 = \mathbb{1}_2, \tag{85a}$$

$$\sigma_j \sigma_k = \begin{cases} +i\sigma_l & \text{if } (j,k,l) \text{ is a cyclic permutation of } (1,2,3), \\ -i\sigma_l & \text{otherwise.} \end{cases} \tag{85b}$$

Example 3.13. Let $p \in [0,1]$, and define a qubit-to-qubit channel $\mathcal{Z}_p\colon M_2(\mathbb{C}) \to M_2(\mathbb{C})$ via

$$\mathcal{Z}_p(\rho) = (1-p)\rho + p\sigma_3\rho\sigma_3. \tag{86}$$

Note that a Kraus representation of this map is given by $\{\sqrt{1-p}\mathbb{1}_A, \sqrt{p}\sigma_3\}$, and hence $\mathcal{Z}_p$ is indeed a quantum channel, usually called *dephasing channel*. Physically, this channel leaves the quantum system intact with probability $1-p$, and afflicts a Z-error with probability p.

Let $\mathcal{P} = \{i^k\sigma_j\colon 0 \le j,k \le 3\} \subset \mathcal{U}(\mathbb{C}^2)$, which is known as the *Pauli group*. It is not hard to check that $\mathcal{P}$ indeed satisfies the properties (G1), (G2), and (G3) above. Furthermore, a straightforward calculation using (85) shows that for all $V \in \mathcal{P}$, we have

$$\mathcal{Z}_p(V \cdot V) = V\mathcal{Z}_p(\cdot)V. \tag{87}$$

Example 3.14. Let $p \in [0,1]$, and define a qubit-to-qubit channel $\mathcal{D}_p\colon M_2(\mathbb{C}) \to M_2(\mathbb{C})$ via

$$\mathcal{D}_p(\rho) = (1-p)\rho + \frac{p}{3}\sum_{j=1}^{3}\sigma_j\rho\sigma_j. \tag{88}$$

This is a quantum channel with Kraus representation $\{\sqrt{1-p}\sigma_0, \sqrt{p/3}\sigma_1, \sqrt{p/3}\sigma_2, \sqrt{p/3}\sigma_3\}$ called the *depolarizing channel*. Physically, this channel leaves the quantum system intact with probability $1-p$, and afflicts an X-, Y-, or Z-error each with equal probability $p/3$. This quantum channel is an important model for the noise affecting a variety of physical quantum systems.

To determine the covariance group of the depolarizing channel, we note that its action on an input state ρ can be rewritten as

$$\mathcal{D}_p(\rho) = (1-q)\rho + q\,\mathrm{tr}(\rho)\frac{1}{2}\mathbb{1}_A, \tag{89}$$

where $q = 4p/3 \in [0,4/3]$. The relation (89) can be checked by direct calculation, and it shows that in fact *every* unitary on A is a symmetry of $\mathcal{D}_p$:

$$\mathcal{D}_p(U \cdot U^\dagger) = U\mathcal{D}_p(\cdot)U^\dagger \qquad \text{for all } U \in \mathcal{U}(\mathbb{C}^2). \tag{90}$$

The next proposition shows that the Choi state once again encodes the covariance properties of a quantum channel, in line with previous characterizations of channels with special properties:

Proposition 3.15. *Let $G \subset \mathcal{U}(A)$ be a group. A channel $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(A)$ is covariant with respect to G if and only if for all $U \in G$,*

$$(\overline{U} \otimes U)\tau^{\Phi}(\overline{U} \otimes U)^{\dagger} = \tau^{\Phi}. \tag{91}$$

Here, $\overline{U}$ denotes the entry-wise complex conjugate of U, taken with respect to the same basis used in the definition of τ^{Φ}.

Proof. Before proving the proposition, we record the following useful fact usually called the *transpose trick*, which can be verified by direct calculation: For any $M \in \mathcal{L}(A)$,

$$(M \otimes \mathbb{1}_A)|\gamma\rangle_{AA} = (\mathbb{1}_A \otimes M^T)|\gamma\rangle_{AA}, \tag{92}$$

where the transposition is taken with respect to the basis used to define $|\gamma\rangle_{AA}$.

Let now $\Phi\colon \mathcal{L}(A) \to \mathcal{L}(A)$ be covariant with respect to G, i.e., for all $U \in G$

$$\Phi(U \cdot U^{\dagger}) = U\Phi(\cdot)U^{\dagger}. \tag{93}$$

By (92), we have

$$(\mathbb{1}_A \otimes U)|\gamma\rangle\langle\gamma|(\mathbb{1}_A \otimes U)^{\dagger} = (U^T \otimes \mathbb{1}_A)|\gamma\rangle\langle\gamma|(U^T \otimes \mathbb{1}_A)^{\dagger}. \tag{94}$$

With $\tau^{\Phi} = (\mathrm{id} \otimes \Phi)(\gamma)$, we obtain

$$\begin{aligned}
(\mathbb{1}_A \otimes U)\tau^{\Phi}(\mathbb{1}_A \otimes U)^{\dagger} &= (\mathbb{1}_A \otimes U)(\mathrm{id} \otimes \Phi)(\gamma)(\mathbb{1}_A \otimes U)^{\dagger} && (95)\\
&= (\mathrm{id}_A \otimes \Phi)((\mathbb{1}_A \otimes U)\gamma(\mathbb{1}_A \otimes U)^{\dagger}) && (96)\\
&= (\mathrm{id}_A \otimes \Phi)((U^T \otimes \mathbb{1}_A)\gamma(U^T \otimes \mathbb{1}_A)^{\dagger}) && (97)\\
&= (U^T \otimes \mathbb{1}_A)(\mathrm{id} \otimes \Phi)(\gamma)(U^T \otimes \mathbb{1}_A)^{\dagger} && (98)\\
&= (U^T \otimes \mathbb{1}_A)\tau^{\Phi}(U^T \otimes \mathbb{1}_A)^{\dagger}, && (99)
\end{aligned}$$

where we used (93) in the second equality, and (94) in the third equality. Conjugating (99) with $\overline{U} \otimes \mathbb{1}_A$ and noting that $\overline{U}U^T = (UU^{\dagger})^T = \mathbb{1}_A^T = \mathbb{1}_A$ now gives (91).

Conversely, we rewrite the relation (91) using the transpose trick (92) and $\overline{U}U^T = \mathbb{1}_A$ to obtain the following, valid for any $U \in G$:

$$(\mathbb{1}_A \otimes U)(\mathrm{id} \otimes \Phi)(\gamma)(\mathbb{1}_A \otimes U)^{\dagger} = (\mathrm{id} \otimes \Phi)((\mathbb{1}_A \otimes U)\gamma(\mathbb{1}_A \otimes U)^{\dagger}). \tag{100}$$

Defining the unitary channel $\mathcal{C}_U(X) := UXU^{\dagger}$ for a given unitary $U \in \mathcal{U}(A)$, the relation (100) says that the channels $\Phi \circ \mathcal{C}_U$ and $\mathcal{C}_U \circ \Phi$ have the same Choi state for all $U \in G$, and hence $\Phi \circ \mathcal{C}_U = \mathcal{C}_U \circ \Phi$ for all $U \in G$ by Theorem 3.3, which means that Φ is covariant with respect to G. □

4 Optimization methods in quantum information theory

In this section we introduce a class of convex optimization problems called *semidefinite programs*. This framework is very useful in quantum information science, since many information-processing tasks can be naturally phrased as semidefinite programs. Consequently, solving the optimization problem gives us information about the quantum information-processing task.

As a motivating example, we consider the fundamental task of *quantum state discrimination*, in which an experimenter is asked to distinguish quantum states drawn from a given set by performing measurements. More precisely, she receives an unknown state from a fixed *quantum state ensemble*, that is, a set of quantum states $\{\rho_1, \dots, \rho_k\}$ together with a probability distribution $(p_1, \dots, p_k)$; with probability p_i the experimenter is given the state ρ_i.

Recall from Section 3.3.1 that a measurement is modeled by a positive operator-valued measure (POVM), a collection of positive semidefinite operators $E_i \geq 0$ satisfying $\sum_i E_i = \mathbb{1}$. For our state discrimination problem we may assume without loss of generality that the POVM consists of k effect operators E_i, with each E_i associated to the state ρ_i. Upon obtaining the outcome i with probability $\operatorname{tr}(E_i\sigma)$ (where σ is one of the states ρ_j for $j = 1, \dots, k$), the experimenter makes the decision that she was given ρ_i. For a fixed POVM $E = \{E_i\}$, elementary probability theory tells us that the *average success probability* of correctly identifying the states from the quantum state ensemble $(\rho_i, p_i)_{i=1}^k$ is given by

$$p_{\text{succ}}(E) = \sum_{i=1}^{k} \Pr(\text{state } \rho_i) \Pr(\text{correctly identifying } i | \text{state } \rho_i) \tag{101}$$

$$= \sum_{i=1}^{k} p_i \operatorname{tr}(E_i\rho_i), \tag{102}$$

where we explicitly denote the dependence of the success probability on the chosen POVM E. Naturally, there are many choices for this POVM, and we are interested in the one maximizing our chance of correctly identifying the state. Mathematically, this amounts to optimizing the success probability over all POVMs,

$$p_{\text{succ}} = \max_{E \text{ POVM}} p_{\text{succ}}(E) \tag{103}$$

$$= \max\left\{\sum_{i=1}^{k} p_i \operatorname{tr}(E_i\rho_i) \colon E_i \geq 0 \text{ for all } i = 1, \dots, k, \sum_{i=1}^{k} E_i = \mathbb{1}\right\}. \tag{104}$$

In (104) we explicitly included the conditions the operators E_i of a POVM have to satisfy. We will see in this section that (104) is an example of a semidefinite program. However, before we can formally introduce these optimization problems, we need to discuss what it means for operators to satisfy semidefinite constraints.

4.1 Operator inequalities

In this section we restrict our attention to Hermitian operators on a Hilbert space:

$$\operatorname{Herm}(\mathcal{H}) = \{X \in \mathcal{L}(\mathcal{H}) \colon X = X^\dagger\}. \tag{105}$$

We can introduce a partial order on $\operatorname{Herm}(\mathcal{H})$ that allows us to compare operators to each other:

Definition 4.1 (Partial order on $\mathrm{Herm}(\mathcal{H})$)**.** For $X, Y \in \mathrm{Herm}(\mathcal{H})$, we say that Y is greater than X, in symbols $Y \geq X$, if $Y - X \geq 0$ is a positive semidefinite operator.

This defines a partial order on $\mathrm{Herm}(\mathcal{H})$ in the following sense: For all $A, B, C \in \mathrm{Herm}(\mathcal{H})$,

- $A \leq A$;
- if $A \leq B$ and $B \leq A$, then $A = B$;
- if $A \leq B$ and $B \leq C$, then also $A \leq C$.

However, $(\mathrm{Herm}(\mathcal{H}), \leq)$ does *not* define a total order: there are Hermitian operators D and E such that neither $D \leq E$ nor $E \leq D$ holds. As an example, consider $D = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $E = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Then neither $D - E = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ nor $E - D = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ is positive, and hence D and E are incomparable with respect to $\leq$.

Remark 4.2. Some authors denote the partial order in Definition 4.1 with a special symbol $\preceq$ in order to distinguish it from the usual 'greater than' relation on the reals. However, we follow the convention of most textbooks and research papers that uses the same symbol for both relations. It will always be clear from context whether we mean the partial order from Definition 4.1 or the usual order on the reals.

The partial order $(\mathrm{Herm}(\mathcal{H}), \leq)$ satisfies a number of useful properties:

Proposition 4.3. *Let $A, B, C, D \in \mathrm{Herm}(\mathcal{H})$.*

(i) *If $A \leq B$ and $C \leq D$, then $A + C \leq B + D$.*

(ii) *If $A \leq B$, then $VAV^\dagger \leq VBV^\dagger$ for all $V \in \mathcal{L}(\mathcal{H})$.*

(iii) *If $A \leq B$, then $\Phi(A) \leq \Phi(B)$ for every positive map $\Phi \colon \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$.*

Proof. (i) We have $B + D - (A + C) = (B - A) + (D - C) \geq 0$ as a sum of positive operators, and hence $B + D \geq A + C$.
(ii) If $B - A \geq 0$ then also $V(B - A)V^\dagger \geq 0$ by the definition of positivity. Expanding gives $VBV^\dagger - VAV^\dagger \geq 0$.
(iii) If $B - A \geq 0$, then also $\Phi(B) - \Phi(A) = \Phi(B - A) \geq 0$ by linearity and positivity of Φ. □

An important consequence of Proposition 4.3 is that we can take convex combinations of operator inequalities to obtain a new one: If $A \leq B$ and $C \leq D$, then also $\lambda A + (1 - \lambda)C \leq \lambda B + (1 - \lambda)D$ for any $\lambda \in [0, 1]$, which follows directly from Proposition 4.3(i) and the fact that $\lambda X \geq 0$ if $\lambda \geq 0$ (as a real number) and $X \geq 0$ (as an operator).

While Proposition 4.3 translates some of the usual properties of the total order $(\mathbb{R}, \leq)$ to Hermitian matrices, there are some important differences. The one we have already mentioned is that the Hermitian matrices are not totally ordered with respect to $\leq$. A second one is that $\leq$ behaves quite differently with respect to multiplication. For example, there are positive operators $A, B \geq 0$ satisfying $A \geq B$, but $A^2 \not\geq B^2$. A simple example are the (2×2)-matrices

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \geq 0 \qquad\qquad B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \geq 0. \tag{106}$$

We have $A \geq B$ since

$$A - B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \geq 0, \tag{107}$$

but the matrix

$$A^2 - B^2 = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix} - \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix} \tag{108}$$

has a negative eigenvalue, and thus $A^2 \not\geq B^2$. These examples show that one has to use caution when working with the partial order $(\mathrm{Herm}(\mathcal{H}), \leq)$.

4.2 Linear and semidefinite programming

We are now ready to introduce semidefinite programs, a special class of convex optimization problems. Our starting point is *linear programs*, in which a linear function is optimized with respect to linear inequality constraints. More formally, let $x \in \mathbb{R}^n$ be a column vector of variables, $c, a_i \in \mathbb{R}^n$ for $i = 1, \ldots, m$ data column vectors, and $b_i \in \mathbb{R}$ for $i = 1, \ldots, m$. Then a general linear program is given by

$$\begin{aligned} &\text{maximize: } c^T x \\ &\text{subject to: } a_i^T x \leq b_i \text{ for } i = 1, \ldots, m. \end{aligned} \tag{109}$$

This general formulation can include different types of constraints. For example, choosing $(a)_j = -\delta_{jk}$ and $b = 0$ leads to a positivity constraint on the k-th component of x, since $a^T x \leq b$ implies $x_k \geq 0$. An equality constraint can be implemented using the observation that $a^T x = b$ if $a^T x \leq b$ and $-a^T x \leq -b$.

Linear programs have a number of useful properties. There are efficient algorithms such as the interior point method to solve linear programs in polynomial time. Furthermore, there is a powerful duality theory for linear programs based on the Lagrangian method. In a nutshell, every linear program of the form (109) (called the *primal* problem) involving a maximization can be turned into a *dual* program involving a minimization instead. The dual has the same optimal solution as the original problem, which is known as *strong duality*. For a detailed discussion of linear programming, we refer to the excellent lecture notes by Weber [19].

Our goal is now to generalize the linear program in (109) to include matrices in the constraints. To this end, let us write out the m constraints $a_i^T x \leq b_i$ in components:

$$\sum_{j=1}^{n} (a_i)_j x_j \leq b_i \quad \text{for } i = 1, \ldots, m, \tag{110}$$

where v_j denotes the j-th component of a vector $v = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$. Defining the vectors

$$\hat{a}_j = \begin{pmatrix} (a_1)_j \\ \vdots \\ (a_m)_j \end{pmatrix} \in \mathbb{R}^m, j = 1, \ldots, n \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m, \tag{111}$$

we may combine the m constraints in (110) into a single vector constraint

$$\sum_{j=1}^{n} x_j \hat{a}_j \leq b, \tag{112}$$

where $\leq$ is defined *component-wise*. The idea is now to replace the vectors $\hat{a}_j, b \in \mathbb{R}^m$ by Hermitian matrices $A_j, B \in \mathrm{Herm}(\mathcal{H})$, and the component-wise inequality $\leq$ by the operator inequality introduced in Definition 4.1. This procedure leads to a constraint of the form

$$\sum_{j=1}^{n} x_j A_j \leq B, \tag{113}$$

and converts the linear program (109) to a so-called *semidefinite* program:

$$\begin{aligned} &\text{maximize: } c^T x \\ &\text{subject to: } \sum_{j=1}^{n} x_j A_j \leq B \\ &\qquad\qquad A_j, B \in \mathrm{Herm}(\mathcal{H}) \text{ for } j = 1, \ldots, n. \end{aligned} \tag{114}$$

Choosing A_j and B to be diagonal matrices turns such a semidefinite program back into an optimization problem as in (109), and hence semidefinite programs generalize linear programs.[9] They share some of the useful properties of linear programs. The interior point method can be generalized to an efficient solving algorithm for semidefinite programs. Moreover, the primal problem (114) can be converted into a dual problem involving a minimization. However, in general we only have *weak duality* in semidefinite programming: the optimal solution of the dual program always bounds the optimal solution of the primal from above, but there might be a gap between the two. For more details on semidefinite programming, we refer to the excellent review by Vandenberghe and Boyd [17] and Section 1.2.3 of the textbook by Watrous [18]. We also mention a recent introductory article on how semidefinite programming arises naturally in quantum information science [16].

To conclude this section, let us make sure that the class of optimization problems (114) is indeed able to encode problems such as the motivating example of quantum state discrimination formulated in (104). For simplicity, we consider the problem of discriminating between two quantum states ρ_1 and ρ_2 with given probabilities p_1 and p_2, respectively. In this case, we are optimizing over the set of two-element POVMs $\{E_1, E_2\}$ such that $E_1, E_2 \geq 0$ and $E_1 + E_2 = \mathbb{1}$. We can eliminate the operator variable E_2 via $E_2 = \mathbb{1} - E_1$, leaving just one variable $E \equiv E_1 \geq 0$. Since $E_2 \geq 0$ by assumption, this gives another constraint $E \leq \mathbb{1}$, leading to the following optimization problem:

$$\begin{aligned} &\text{maximize: } p_1 \operatorname{tr}(E\rho_1) + p_2 \operatorname{tr}((\mathbb{1} - E)\rho_2) \\ &\text{subject to: } E \in \mathrm{Herm}(\mathcal{H}), \\ &\qquad\qquad 0 \leq E \leq \mathbb{1}. \end{aligned} \tag{115}$$

$\mathrm{Herm}(\mathcal{H})$ has a real vector space structure, and hence there is a basis $\{\sigma_i\}$ of Hermitian operators such that $E = \sum_i x_i \sigma_i$ for some $x_i \in \mathbb{R}$. The two constraints $E \geq 0$ and $E \leq \mathbb{1}$ then translate to

$$\sum_i x_i (-\sigma_i) \leq 0 \tag{116}$$

$$\sum_i x_i \sigma_i \leq \mathbb{1}, \tag{117}$$

[9] For real diagonal matrices, the partial order $\leq$ from Definition 4.1 reduces to a component-wise inequality for the diagonal elements, similar to (112). For example, considering diagonal (2×2)-matrices $X = \left(\begin{smallmatrix} x_1 & 0 \\ 0 & x_2 \end{smallmatrix}\right)$ and $Y = \left(\begin{smallmatrix} y_1 & 0 \\ 0 & y_2 \end{smallmatrix}\right)$, we have $X \leq Y$ as an operator inequality if and only if $x_1 \leq y_1$ and $x_2 \leq y_2$ as real inequalities.

which may be combined into a single relation using block-diagonal matrices as

$$\sum_i x_i \begin{pmatrix} -\sigma_i & 0 \\ 0 & \sigma_i \end{pmatrix} \leq \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{1} \end{pmatrix}. \tag{118}$$

This constraint is exactly of the form as in (113). For the objective function,

$$\begin{aligned} p_1 \operatorname{tr}(E\rho_1) + p_2 \operatorname{tr}((\mathbb{1} - E)\rho_2) &= \operatorname{tr}(E(p_1\rho_1 - p_2\rho_2)) + p_2 && (119) \\ &= \sum_i x_i \operatorname{tr}(\sigma_i(p_1\rho_1 - p_2\rho_2)) + p_2 && (120) \\ &= c^T x + p_2, && (121) \end{aligned}$$

where we set $c_i = \operatorname{tr}(\sigma_i(p_1\rho_1 - p_2\rho_2))$. Since the objective function is now also in the form of (114) (modulo the constant term p_2, which can be safely ignored in the maximization), we have successfully turned the problem (115) into the form of a general semidefinite program in (114).

Note that this conversion procedure is just for illustrative purposes, and generally *not* necessary for solving a semidefinite program or finding the dual problem. For example, the state discrimination problem (104) can directly be entered in (and solved by) `cvx` [7, 8], a package for specifying and solving convex programs available for `MATLAB` and `python`.

References

[1] John S. Bell. "On the Einstein Podolsky Rosen paradox". *Physics Physique Fizika* 1.3 (1964), p. 195.

[2] Charles H. Bennett and Gilles Brassard. "Quantum cryptography: Public key distribution and coin tossing". *1984 IEEE International Conference on Computers, Systems and Signal Processing*. Bangalore, India, 1984, pp. 175–179.

[3] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels". *Physical Review Letters* 70.13 (1993), pp. 1895–1899.

[4] Albert Einstein, Boris Podolsky, and Nathan Rosen. "Can quantum-mechanical description of physical reality be considered complete?" *Physical Review* 47.10 (1935), p. 777.

[5] William Fulton and Joe Harris. *Representation Theory: A First Course*. Graduate Texts in Mathematics. Springer New York, 1991.

[6] Sevag Gharibian. "Strong NP-hardness of the quantum separability problem". *Quantum Information & Computation* 10.3&4 (2010), pp. 343–360. arXiv: 0810.4507 [quant-ph].

[7] Michael Grant and Stephen Boyd. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. http://cvxr.com/cvx. Mar. 2014.

[8] Michael Grant and Stephen Boyd. "Graph implementations for nonsmooth convex programs". *Recent Advances in Learning and Control*. Ed. by V. Blondel, S. Boyd, and H. Kimura. Lecture Notes in Control and Information Sciences. http://stanford.edu/~boyd/graph_dcp.html. Springer-Verlag Limited, 2008, pp. 95–110.

[9] Leonid Gurvits. "Classical Deterministic Complexity of Edmonds' Problem and Quantum Entanglement". *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*. STOC '03. San Diego, CA, USA: Association for Computing Machinery, 2003, 10–19. arXiv: quant-ph/0303055.

[10] Karol Horodecki, Michał Horodecki, Paweł Horodecki, and Jonathan Oppenheim. "General Paradigm for Distilling Classical Key From Quantum States". *IEEE Transactions on Information Theory* 55.4 (2009), pp. 1898–1929. arXiv: quant-ph/0506189.

[11] Karol Horodecki, Michał Horodecki, Paweł Horodecki, and Jonathan Oppenheim. "Secure Key from Bound Entanglement". *Physical Review Letters* 94 (16 2005), p. 160502. arXiv: quant-ph/0309110.

[12] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. "Mixed-State Entanglement and Distillation: Is there a "Bound" Entanglement in Nature?" *Physical Review Letters* 80.24 (1998), pp. 5239–5242. quant-ph: quant-ph/9801069.

[13] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. "Separability of mixed states: necessary and sufficient conditions". *Physics Letters A* 223.1 (1996), pp. 1–8. arXiv: quant-ph/9605038.

[14] Paweł Horodecki, Michał Horodecki, and Ryszard Horodecki. "Binding entanglement channels". *Journal of Modern Optics* 47.2-3 (2000), pp. 347–354. arXiv: quant-ph/9904092.

[15] John F. Humphreys. *A Course in Group Theory*. Oxford science publications. Oxford University Press, 1996.

[16] Vikesh Siddhu and Sridhar Tayur. "Five Starter Pieces: Quantum Information Science via Semidefinite Programs". *arXiv preprint* (2021). arXiv: 2112.08276 [quant-ph].

[17] Lieven Vandenberghe and Stephen Boyd. "Semidefinite programming". *SIAM Review* 38.1 (1996). Available online, pp. 49–95.

[18] John Watrous. *The Theory of Quantum Information*. Available online. Cambridge University Press, 2018.

[19] Richard Weber. *Lecture notes on Optimization*. Available online. University of Cambridge, 2010.

[20] Mark M. Wilde. *Quantum information theory*. Cambridge University Press, 2013. arXiv: 1106.1445 [quant-ph].