Part III

Probabilistic Quantum Information

Chapter 6

The Density-Operator Formalism

6.1 Introduction

We introduce here the density-operator formalism, where quantum states are described by operators (of a certain form), rather than by vectors as in the state-vector formalism introduced in Chapter 1 and considered so far. To motivate the density operator formalism, we illustrate a couple of limitations of the state vector formalism.

The first one is in the context of *randomized* states. Consider a quantum system with a state that is given by a certain state vector with a certain probability. This could be in the context of an experiment where a quantum system is prepared in a way that depends on the outcome of a classical experiment, e.g. a coin toss. Or, we want to consider the postmeasurement state when we do not have access to the classical measurement outcome. How should we capture such a randomized state? The straightforward way is by means of a probability distribution over state vectors, i.e., by a list $\{(\varepsilon_1, |\varphi_1\rangle), \ldots, (\varepsilon_n, |\varphi_n\rangle)\}$ that indicates that the the state of the considered quantum system is $|\varphi_\ell\rangle$ with probability ε_ℓ. The caveat with this representation is that it is not unique. For example, it is not too hard to see that the two randomized states respectively described by $\{(\frac{1}{2}, |0\rangle), (\frac{1}{2}, |1\rangle)\}$ and $\{(\frac{1}{2}, |+\rangle), (\frac{1}{2}, |-\rangle)\}$ behave *identically* under any quantum operation, and thus they represent *the same* (randomized) quantum state. For instance, in both cases, a measurement in the computational basis produces a random bit as outcome.

The other limitation of the state vector formalism is that it does not allow us (in an obvious way) to express the state of a subsystem of an bipartite system. For example, what is the state of subsystem A given that the state of AB is $|\Phi\rangle = (|0\rangle|0\rangle + |1\rangle|1\rangle)/\sqrt{2}$? Note that it is not $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$: measuring one qubit of an EPR pair in the Hadamard basis produced a random measurement outcome, measuring $|+\rangle$ produces a deterministic outcome.

The density-operator formalism allows us to conveniently deal with the above issues. Given that in the density-operator formalism quantum states are described by operators, actions on quantum states are then described by superoperators. An additionaly nice property is that we will find a unifying notion for the possible actions that can be applied to a quantum system — in contrast to the state-vector formalism, where measurements and unitary operations appear to be two very different kind of actions, captured by different mathematical objects.

6.2 Density Operators

Let $\mathcal{H}$ be a Hilbert space. Recall that in Section 1.7, we have already seen that the map

$$\mathcal{S}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}), |\varphi\rangle \mapsto |\varphi\rangle\langle\varphi|$$

is an injection modulo the equivalence relation $\equiv$, which identifies state vectors that are equal up to the phase, and thus behave identically as quantum states. Thus, $|\varphi\rangle\langle\varphi|$ is a *faithful* representation of the state $|\varphi\rangle$, and it avoids the ambiguity caused by the phase.

Consider now a *randomized* state, given by a list $\{(\varepsilon_1, |\varphi_1\rangle), \ldots, (\varepsilon_L, |\varphi_L\rangle)\}$ as discussed above, where $|\varphi_1\rangle, \ldots, |\varphi_L\rangle \in \mathcal{S}(\mathcal{H})$ and $\varepsilon_1, \ldots, \varepsilon_L \geq 0$ with $\sum_\ell \varepsilon_\ell = 1$, with the understanding that the state of the considered quantum system is $|\varphi_\ell\rangle$ with probability ε_ℓ. Furthermore, let $\mathbf{M} = \{M_i\}_{i\in I}$ be a measurement in $\mathcal{M}eas_I(\mathcal{H})$. By Born's rule, *conditioned* on the state being $|\varphi_\ell\rangle$ for a particular ℓ, the probability to observe outcome $i \in I$ is $p_{i|\ell} := \langle\varphi_\ell|M_i^\dagger M_i|\varphi_i\rangle$. Hence, by basic probability theory, the (average) probability to observe $i \in I$ is

$$p_i = \sum_{\ell=1}^{n} \varepsilon_\ell \, p_{i|\ell} = \sum_{\ell=1}^{L} \varepsilon_\ell \, \langle\varphi_\ell|M_i^\dagger M_i|\varphi_i\rangle = \sum_{\ell=1}^{n} \varepsilon_\ell \operatorname{tr}\bigl(M_i^\dagger M_i|\varphi_\ell\rangle\langle\varphi_\ell|\bigr) = \operatorname{tr}\Biggl(M_i^\dagger M_i \sum_{\ell=1}^{n} \varepsilon_\ell|\varphi_\ell\rangle\langle\varphi_\ell|\Biggr) .$$

This suggests to introduce

$$\rho := \sum_{\ell=1}^{L} \varepsilon_\ell|\varphi_\ell\rangle\langle\varphi_\ell|$$

as a description of the randomized state. By the properties

$$\langle\psi|\rho|\psi\rangle = \sum_\ell \varepsilon_\ell \, \langle\psi||\varphi_\ell\rangle\langle\varphi_\ell||\psi\rangle = \sum_\ell \varepsilon_\ell \, |\langle\varphi_\ell|\psi\rangle|^2 \geq 0 \qquad \forall\, |\psi\rangle \in \mathcal{H}$$

and

$$\operatorname{tr}(\rho) = \sum_\ell \varepsilon_\ell \operatorname{tr}(|\varphi_\ell\rangle\langle\varphi_\ell|) = \sum_\ell \varepsilon_\ell = 1$$

this motivates the following.

Definition 6.1. *An operator $\rho \in \mathcal{L}(\mathcal{H})$ is called a* **density operator** *(or* **matrix***) if $\rho \geq 0$ and $\operatorname{tr}(\rho) = 1$. $\mathcal{D}(\mathcal{H})$ denotes the set of density operators in $\mathcal{L}(\mathcal{H})$.*

We have seen that every (possibly randomized) state gives rise to a density operator. Vice versa, every density operator represents a (possibly randomized) state. Indeed, it follows from spectral decomposition (Theorem 0.3) that every $\rho \in \mathcal{D}(\mathcal{H})$ decomposes into $\rho = \sum_\ell \varepsilon_\ell|\varphi_\ell\rangle\langle\varphi_\ell|$ for orthonormal vectors $|\varphi_1\rangle, \ldots, |\varphi_d\rangle$, where $d = \dim(\mathcal{H})$. From positivity it then follows that $\varepsilon_\ell = \langle\varphi_\ell|\rho|\varphi_\ell\rangle \geq 0$, and from the trace condition and noting that $\operatorname{tr}(|\varphi_\ell\rangle\langle\varphi_\ell|) = \langle\varphi_\ell|\varphi_\ell\rangle = 1$, we get that $\sum_\ell \varepsilon_\ell = \operatorname{tr}(\rho) = 1$.

A special role play the density operators that correspond to a (deterministic) state vector.

Definition 6.2. *A density operator $\rho \in \mathcal{D}(\mathcal{H})$ is said to be* **pure** *if its rank is $\operatorname{rank}(\rho) = 1$. A density operator is called* **mixed** *if it is not (necessarily) pure.*

By elementary properties, $\rho \in \mathcal{D}(\mathcal{H})$ is pure if and only if there exists $|\varphi\rangle \in \mathcal{S}(\mathcal{H})$ with

$$\rho = |\varphi\rangle\langle\varphi| .$$

Indeed, having rank 1 and being Hermitian implies that $\rho = |\varphi\rangle\langle\varphi|$ for some vector $|\varphi\rangle \in \mathcal{H}$ (see the corresponding discussion in Section 0.2), and the requirement on the trace them implies that $\||\varphi\rangle\|^2 = \langle\varphi|\varphi\rangle = \operatorname{tr}(|\varphi\rangle\langle\varphi|) = \operatorname{tr}(\rho) = 1$.

As should be clear from the above, in the density-operator formalism Born's rule becomes

$$p_i = \operatorname{tr}(M_i^\dagger M_i \rho) = \operatorname{tr}(M_i \rho M_i^\dagger) \qquad \text{and} \qquad \rho^i = \frac{1}{p_i} M_i \rho M_i^\dagger .$$

Also, a unitary $U \in \mathcal{U}(\mathcal{H})$ acts on a density operator $\rho \in \mathcal{D}(\mathcal{H})$ as

$$\rho \mapsto U \rho U^\dagger \,.$$

This obviously extends to isometries. In particular, the isometry $\mathbb{I}_A \otimes |0\rangle_B : \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_B$, $|\varphi\rangle_A \mapsto |\varphi\rangle_A \otimes |0\rangle_B$, which adds an ancilla to a given state, acts as

$$\rho_A \mapsto \rho_A \otimes |0\rangle\langle 0|_B \,.$$

Thus, "quantum operations" are now described by *superoperators* acting on $\mathcal{L}(\mathcal{H})$. The following convention on the notation will be useful.

Remark 6.1. For any $L \in \mathcal{L}(\mathcal{H}, \mathcal{H}')$, by abuse of notation we identify the *operator* L with the *superoperator* $L \in \mathcal{L}(\mathcal{L}(\mathcal{H}), \mathcal{L}(\mathcal{H}'))$ given by

$$L(R) := LRL^\dagger \,.$$

6.3 Partial Trace

Consider a density operator $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, describing the state of a bipartite quantum system AB. How can one then describe the subsystem B when considered as a stand-alone system? We answer this question here.

Definition 6.3. *For any Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the* **partial trace** tr_A *is the superoperator*

$$\mathrm{tr}_A := \mathrm{tr} \otimes id_B : \mathcal{L}(AB) \to \mathbb{C} \otimes \mathcal{L}(B) = \mathcal{L}(B) \,,$$

where $\mathrm{tr} : \mathcal{L}(A) \to \mathbb{C}$ *is the standard trace on $\mathcal{L}(A)$.* tr_B *is defined accordingly.*

The following is a useful alternative characterization.

Proposition 6.1. *The partial trace tr_A coincides with the adjoint (w.r.t. the Hilbert-Schmidt inner product) of the superoperator $\mathcal{L}(B) \to \mathcal{L}(AB)$, $L \mapsto \mathbb{I}_A \otimes L$; in other words, it is the unique superoperator $\mathrm{tr}_A \colon \mathcal{L}(AB) \to \mathcal{L}(B)$ that satisfies*

$$\mathrm{tr}\big(L \cdot \mathrm{tr}_A(R)\big) = \mathrm{tr}\big((\mathbb{I}_A \otimes L)R\big)$$

for all $L \in \mathcal{L}(B)$ and $R \in \mathcal{L}(AB)$.

Proof. For every $L_B \in \mathcal{L}(B)$ and for $R_{AB} \in \mathcal{L}(AB)$ of the form $R_{AB} = R_A \otimes R_B$, we have

$$\mathrm{tr}\big(L_B \cdot \mathrm{tr}_A(R_{AB})\big) = \mathrm{tr}(R_A) \cdot \mathrm{tr}(L_B \cdot R_B) = \mathrm{tr}\big(R_A \otimes (L_B \cdot R_B)\big) = \mathrm{tr}\big((\mathbb{I}_A \otimes L_B) \cdot R_{AB}\big) \,,$$

and by linearity it follows that the equality holds for *every* $R_{AB} \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$. □

Note that Proposition 6.1 in particular implies that $\mathrm{tr} \circ \mathrm{tr}_A(R) = \mathrm{tr}(R)$ for any $R \in \mathcal{L}(AB)$, i.e., the partial trace preserves the (standard) trace. It can also be used to show that the partial trace preserves positivity.

Lemma 6.2. *For any $R \in \mathcal{L}(AB)$: if $R \geq 0$ then $\mathrm{tr}_A(R) \geq 0$.*

Proof. For any $|\varphi\rangle \in \mathcal{H}_B$ we have

$$\begin{aligned}\langle\varphi|\mathrm{tr}_A(R)|\varphi\rangle = \mathrm{tr}\big(|\varphi\rangle\langle\varphi|\mathrm{tr}_A(R)\big) = \mathrm{tr}\big((\mathbb{I}_A \otimes |\varphi\rangle\langle\varphi|)R\big) = \sum_i \mathrm{tr}\big((|i\rangle\langle i| \otimes |\varphi\rangle\langle\varphi|)R\big) \\ = \sum_i \mathrm{tr}\big(|i\rangle|\varphi\rangle\langle i|\langle\varphi|R\big) = \sum_i \langle i|\langle\varphi|R|i\rangle|\varphi\rangle \geq 0\end{aligned}$$

where $\{|i\rangle\}_{i \in I}$ is an arbitrary orthonormal basis of $\mathcal{H}_A$. □

It thus follows that the partial trace tr_A maps a density operator $\rho_{AB} \in \mathcal{D}(AB)$ into a density operator $\rho_B = \mathrm{tr}_A(\rho_{AB}) \in \mathcal{D}(B)$, called the **reduced** density operator. We also say that we **trace out** system A as synonym for applying tr_A. When ρ_{AB} is clear from the context, we may simply write ρ_B for $\mathrm{tr}_A(\rho_{AB})$. Proposition 6.1 ensures that the reduced density operator is indeed the proper description of the state of a subsystem, in that it uniquely determines the behavior of the subsystem under any measurement.

We conclude with working out the reduced density operator of a pure state. If $\{|i\rangle\}_{i\in I}$ is an orthonormal basis of $\mathcal{H}_A$ and $|\Omega\rangle = \sum_i \alpha_i |i\rangle|\psi_i\rangle \in \mathcal{S}(AB)$ then

$$|\Omega\rangle\langle\Omega| = \sum_{ij} \bar{\alpha}_i \alpha_j |i\rangle\langle j| \otimes |\psi_i\rangle\langle\psi_j|$$

and thus

$$\mathrm{tr}_A(|\Omega\rangle\langle\Omega|) = \sum_{ij} \bar{\alpha}_i \alpha_j \, \mathrm{tr}_A(|i\rangle\langle j| \otimes |\psi_i\rangle\langle\psi_j|) = \sum_i |\alpha_i|^2 |\psi_i\rangle\langle\psi_i| \, .$$

6.4 Purification

As we have just seen, tracing out part of a *pure* state typically leads to a *mixed* state. We now look at the question if every mixed state can be obtained by tracing out part of a pure state, and we answer it in the affirmative. Such a pure state is called a **purification** of the given mixed state.

Theorem 6.3 (Purification theorem)**.** *Let $\rho_B \in \mathcal{D}(\mathcal{H}_B)$ be an arbitrary density operator. Then there exists a state vector $|\varphi\rangle \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, where $\mathcal{H}_A = \mathcal{H}_B$, so that $\rho_B = \mathrm{tr}_A(|\varphi\rangle\langle\varphi|)$.*

Proof. By spectral decomposition, ρ_B can be written as $\rho_B = \sum_{\ell=1}^d \varepsilon_\ell |\psi_\ell\rangle\langle\psi_\ell|$ with $d = \dim(\mathcal{H}_B)$. Let $\{|1\rangle, \dots, |d\rangle\}$ be an orthonormal basis of $\mathcal{H}_A = \mathcal{H}_B$, and consider the pure state

$$|\varphi\rangle = \sum_\ell \sqrt{\varepsilon_\ell}\, |\ell\rangle|\psi_\ell\rangle \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B) \, .$$

Tracing out A yields

$$\mathrm{tr}_A(|\varphi\rangle\langle\varphi|) = \sum_{\ell,m} \sqrt{\varepsilon_\ell}\sqrt{\varepsilon_m}\, \mathrm{tr}_B\big(|\ell\rangle\langle m| \otimes |\psi_\ell\rangle\langle\psi_m|\big) = \sum_{\ell,m} \sqrt{\varepsilon_\ell \varepsilon_m}\, \langle m|\ell\rangle |\psi_\ell\rangle\langle\psi_m| = \sum_{\ell=1}^d \varepsilon_\ell |\psi_\ell\rangle\langle\psi_\ell| \, ,$$

which is ρ_B. □

By the freedom of the choice of the basis, it follows immediately that a purification is not unique: different purifications can be obtained by acting with a unitary on the purifying system. The theorem below shows that this is the only freedom in the purification.

Theorem 6.4 (Uniqueness of purification)**.** *Let two state vectors $|\varphi\rangle, |\varphi'\rangle \in \mathcal{S}(AB)$ be so that $\mathrm{tr}_A(|\varphi\rangle\langle\varphi|) = \mathrm{tr}_A(|\varphi'\rangle\langle\varphi'|)$. Then, there exists $U \in \mathcal{U}(A)$ such that $|\varphi\rangle = (U \otimes \mathbb{I}_B)|\varphi'\rangle$.*

Remark 6.2. It is easy to see that the above extends to purifications $|\varphi\rangle$ and $|\varphi'\rangle$ that have different respective "purifying systems" $\mathcal{H}_A$. The unitary U then gets replaced by an isometry that maps the "smaller" purification into the "bigger" one (in terms of dimension).

Proof. Let $\rho_B = \mathrm{tr}_A(|\varphi\rangle\langle\varphi|) = \mathrm{tr}_A(|\varphi'\rangle\langle\varphi'|) \in \mathcal{D}(B)$. By spectral decomposition, we can write $\rho_B = \sum_i \lambda_i |i\rangle\langle i|$ for some orthonormal basis $\{|i\rangle\}_{i\in I}$ of $\mathcal{H}_B$. We can now write $|\varphi\rangle = \sum_i \alpha_i |\varphi_i\rangle|i\rangle$ where $\sum_i |\alpha_i|^2 = 1$ and $|\varphi_i\rangle \in \mathcal{S}(A)$ for all $i \in I$. Furthermore, by adjusting the $|\varphi_i\rangle$'s appropriately, we may assume that $\alpha_i \geq 0$ for all $i \in I$. From

$$\sum_i \lambda_i |i\rangle\langle i| = \mathrm{tr}_A(|\varphi\rangle\langle\varphi|) = \sum_{i,j} \alpha_i \alpha_j \langle\varphi_j|\varphi_i\rangle |i\rangle\langle j|$$

it follows that $\alpha_i^2 = \lambda_i$ for all $i \in I$, and $\alpha_i\alpha_j\langle\varphi_j|\varphi_i\rangle = 0$ for all $i \neq j \in I$. Therefore, for $i \neq j$, $\langle\varphi_j|\varphi_i\rangle = 0$ unless $\alpha_i = 0$ or $\alpha_j = 0$, and thus the corresponding λ_i or λ_j vanishes as well. Hence, $|\varphi\rangle = \sum_i \sqrt{\lambda_i}|\varphi_i\rangle|i\rangle$, where the $|\varphi_i\rangle$'s with $\lambda_i \neq 0$ are orthonormal. The same holds for $|\varphi'\rangle$, namely $|\varphi'\rangle = \sum_i \sqrt{\lambda_i}|\varphi'_i\rangle|i\rangle$, where the $|\varphi'_i\rangle$'s with $\lambda_i \neq 0$ are orthonormal. It thus follows that indeed $|\varphi\rangle = (U \otimes \mathbb{I}_B)|\varphi'\rangle$, where $U \in \mathcal{U}(A)$ is a unitary that satisfies $|\varphi_i\rangle = U|\varphi'_i\rangle$ for every $i \in I$ with $\lambda_i \neq 0$; the existence of such a U follows from the orthonormality. □

From this proof, we can also extract the following result.

Theorem 6.5 (Schmidt decomposition)**.** *For any $|\varphi\rangle \in \mathcal{S}(AB)$ there exist respective orthonormal bases $\{|e_1\rangle, \ldots, |e_{d_A}\rangle\}$ and $\{|f_1\rangle, \ldots, |f_{d_B}\rangle\}$ of $\mathcal{H}_A$ and $\mathcal{H}_B$ such that*

$$|\varphi\rangle = \sum_{i=1}^{d_{min}} \mu_i |e_i\rangle|f_i\rangle\,,$$

where $d_{min} = \min\{d_A, d_B\}$, and $\mu_1, \ldots, \mu_{d_{min}} \geq 0$ and $\sum_i \mu_i^2 = 1$.

What is crucial is that both, the $|e_i\rangle$'s as well as the $|f_i\rangle$'s are mutually orthogonal, and there are no "cross products" $|e_i\rangle|f_j\rangle$.

6.5 Incorporating Classical Information

As discussed in Section 2.4, "classical information" is captured by an element x that is selected from a finite non-empty set $\mathcal{X}$, and it can be embedded into "quantum information" by fixing an orthonormal basis $\{|x\rangle\}_{x\in\mathcal{X}}$ of $\mathcal{H} = \mathbb{C}^{|\mathcal{X}|}$ and by associating $x \in \mathcal{X}$ with the state $|x\rangle \in \mathcal{S}(\mathcal{H})$, respectively $|x\rangle\langle x| \in \mathcal{D}(\mathcal{H})$ when using the density-operator formalism.

Given that we can now deal with *randomized* states, it is now natural to also consider *randomized* "classical information", which is formally captured by a *random variable* X with range $\mathcal{X}$, where here and most of the time, we leave the (always finite) probability space implicit and understand X to be described by its distribution $P_X : \mathcal{X} \to [0,1]$. This motivates the following.

Definition 6.4. *Let X be a random variable with finite range $\mathcal{X}$, and let $P_X : \mathcal{X} \to [0,1]$ be its distribution. Furthermore, let $\{|x\rangle\}_{x\in\mathcal{X}}$ be an orthonormal basis of $\mathcal{H}_X = \mathbb{C}^{|\mathcal{X}|}$. Then,*

$$\rho_X = \sum_x P_X(x)|x\rangle\langle x| \in \mathcal{D}(\mathcal{H}_X)$$

is called the **density operator representation** *of P_X (or of X) w.r.t. $\{|x\rangle\}_{x\in\mathcal{X}}$.*

Vice versa, an arbitrary density operator $\rho_X \in \mathcal{D}(\mathcal{H}_X)$ is called **classical** *w.r.t. $\{|x\rangle\}_{x\in\mathcal{X}}$ if it is of the above form for some distribution P_X. We express this by writing $\rho_X \in \mathcal{D}(\mathcal{X})$.*

We often leave the basis $\{|x\rangle\}_{x\in\mathcal{X}}$ implicit, but take it as understood that the statements are with respect to some fixed choice of $\{|x\rangle\}_{x\in\mathcal{X}}$.

Note that the density operator representation of the *uniform distribution* over $\mathcal{X}$ is given by

$$\frac{1}{|\mathcal{X}|}\sum_x |x\rangle\langle x| = \frac{1}{|\mathcal{X}|}\mathbb{I}_X \in \mathcal{D}(\mathcal{X})\,.$$

This is also called the **completely** (or **fully**) **mixed state** and denoted by μ_X.

We also want to capture a situation where (randomized) classical information X is correlated with quantum information, in the sense that the state of a system E depends on the value that X takes on, i.e., its state is given by $\rho_{\mathsf{E}|X=x}$ if $X = x$. For instance, think of an "experimenter" that first tosses a coin and then prepares a state depending on the outcome of the coin toss, or think of the post-measurement state that depends on the classical measurement outcome.

This is done as follows.

Definition 6.5. *Let $\mathcal{H}_X$ and $\mathcal{H}_\mathsf{E}$ be Hilbert spaces and $\{|x\rangle\}_{x\in\mathcal{X}}$ an orthonormal basis on $\mathcal{H}_X$. A density operator $\rho_{X\mathsf{E}} \in \mathcal{D}(\mathcal{H}_X \otimes \mathcal{H}_\mathsf{E})$ is said to be a* **hybrid state** *with classical X in $\mathcal{X}$ (w.r.t. $\{|x\rangle\}_{x\in\mathcal{X}}$) if it is of the form*

$$\rho_{X\mathsf{E}} = \sum_x P_X(x)|x\rangle\langle x| \otimes \rho_{\mathsf{E}|X=x}$$

for a distribution P_X and density operator $\rho_{\mathsf{E}|X=x} \in \mathcal{D}(\mathcal{H}_\mathsf{E})$ for every $x \in \mathcal{X}$. We express this by writing $\rho_{X\mathsf{E}} \in \mathcal{D}(\mathcal{X} \otimes \mathcal{H}_\mathsf{E})$.

Recall that the state of system E alone is then described by the reduced density operator

$$\rho_\mathsf{E} = \mathrm{tr}_X(\rho_{X\mathsf{E}}) = \sum_x P_X(x)\rho_{\mathsf{E}|X=x}\,.$$

Furthermore, if $\rho_{X\mathsf{E}} \in \mathcal{D}(\mathcal{X} \otimes \mathcal{H}_\mathsf{E})$ is a hybrid state with classical X in $\mathcal{X}$ (and as such we can also understand X as a random variable), and if $\lambda : \mathcal{X} \to \{\texttt{true}, \texttt{false}\}$ is a predicate on $\mathcal{X}$, then we can consider the *event* $\Lambda = \lambda(X)$ and the states[1]

$$\rho_{X\mathsf{E}|\Lambda} = \sum_x P_{X|\Lambda}(x)|x\rangle\langle x| \otimes \rho_{\mathsf{E}|X=x} \qquad \text{and} \qquad \rho_{\mathsf{E}|\Lambda} = \mathrm{tr}_X(\rho_{X\mathsf{E}|\Lambda}) = \sum_x P_{X|\Lambda}(x)\rho_{\mathsf{E}|X=x}\,,$$

where $P_{X|\Lambda}$ is the naturally given conditional probability distribution of X conditioned on Λ.[2] These two density operators describe the joint and the single quantum system *given that Λ occurs*; we sometimes refer to such density operators as **conditional** states. Finally, if $\rho_{X\mathsf{E}}$ in $\mathcal{D}(\mathcal{X} \otimes \mathcal{H}_\mathsf{E})$ and $f : \mathcal{X} \to \mathcal{Y}$ is a function then $\rho_{Xf(X)\mathsf{E}}$ is naturally understood as

$$\rho_{Xf(X)\mathsf{E}} = \sum_x P_X(x)|x\rangle\langle x| \otimes |f(x)\rangle\langle f(x)| \otimes \rho_{\mathsf{E}|X=x}\,,$$

where $\{|y\rangle\}_{y\in\mathcal{Y}}$ is some fixed orthonormal basis of $\mathcal{H}_Y = \mathbb{C}^{|\mathcal{Y}|}$, and therefore

$$\rho_{f(X)\mathsf{E}} = \sum_x P_X(x)|f(x)\rangle\langle f(x)| \otimes \rho_{\mathsf{E}|X=x} = \sum_y P_{f(X)}(y)|y\rangle\langle y| \otimes \rho_{\mathsf{E}|f(X)=y}\,,$$

as can easily be verified.

[1]Note that this notation is consistent when considering the event $\Lambda = [X = x]$.

[2]I.e., $P_{X|\Lambda}(x) = P_X(x)/P[\Lambda]$ if $\lambda(x)$, where $P[\Lambda] = \sum_{x:\lambda(x)} P_X(x)$, and $P_{X|\Lambda}(x) = 0$ otherwise.

By treating pairs (and triples etc.) of random variables as a single random variable, this formalism naturally extends to states that depend on several, possibly dependent, random variables X, Y etc. To simplify notation, we often write ρ_E^x instead of $\rho_{E|X=x}$.[3]

We quickly discuss a concrete but important example. Let $|\varphi\rangle\langle\varphi| \in \mathcal{D}(\mathcal{H}_E)$ be a density operator (which we assume to be pure for simplicity), and consider a measurement $\mathbf{M} = \{M_x\}_{x\in\mathcal{X}}$ in $\mathcal{M}eas_{\mathcal{X}}(\mathcal{H}_E)$. By Born's rule, the result of the measurement consists of a classical measurement outcome $x \in \mathcal{X}$, occurring with probability $P_X(x) = p_x = \langle\varphi|M_x^\dagger M_x|\varphi\rangle = \mathrm{tr}\big(M_x^\dagger M_x|\varphi\rangle\langle\varphi|\big)$, and of the post-measurement state $\rho_{E|X=x} = |\varphi^x\rangle\langle\varphi^x|$ given by $|\varphi^x\rangle = M_x|\varphi\rangle/\sqrt{p_x}$. The above notation allows us now to very cleanly and compactly write the action of a measurement in terms of a superoperator that maps $|\varphi\rangle\langle\varphi| \in \mathcal{D}(\mathcal{H}_E)$ to

$$\rho_{XE} = \sum_x P_X(x)|x\rangle\langle x| \otimes \rho_{E|X=x} = \sum_x |x\rangle\langle x| \otimes M_x|\varphi\rangle\langle\varphi|M_x^\dagger = \sum_x \big(|x\rangle\otimes M_x\big)|\varphi\rangle\langle\varphi|\big(\langle x|\otimes M_x^\dagger\big)$$

in $\mathcal{D}(\mathcal{X} \otimes \mathcal{H}_E)$.

6.6 General Quantum Operations

Let $\mathfrak{T}$ be a superoperator $\mathcal{L}(A) \to \mathcal{L}(A')$, which we may also write as $\mathfrak{T}_A$ (or as $\mathfrak{T}_{A\to A'}$) in order to emphasize its doman (and range). In order for such a superoperator to be a meaningful *quantum operation*, we want that it maps density operators to density operators. Thus, it should be **positive** and **trace-preserving**. However, it is also possible to act on a subsystem of a bipartite system, and thus we actually need to require that $\mathfrak{T}_A(\rho_{AE}) \in \mathcal{D}(A'E)$ for any Hilbert space $\mathcal{H}_E$ and any $\rho_{AE} \in \mathcal{D}(AE)$, where $\mathfrak{T}_A(\rho_{AE})$ is understood as $(\mathfrak{T}_A \otimes id_E)(\rho_{AE})$. This motivates the following.

Definition 6.6. *A superoperator* $\mathfrak{T} : \mathcal{L}(A) \to \mathcal{L}(A')$ *is a* **CPTP map**, *if it is* **completely positive**, *meaning that for any Hilbert space* $\mathcal{H}_E$ *and any* $R \in \mathcal{L}(AE)$*:*

$$R_{AE} \geq 0 \;\Rightarrow\; \mathfrak{T}_A(R_{AE}) \geq 0\,,$$

and it is **trace preserving**, *meaning that for all* $R \in \mathcal{L}(A)$*:* $\mathrm{tr} \circ \mathfrak{T}_A(R) = \mathrm{tr}(R)$.

The canonical example of a superoperator that is positive but not completely positive is the **transposition map**, as introduced in Definition 0.3. To show that the transpose is not completely positive, observe that for $|\Omega\rangle := \sum_i |i\rangle|i\rangle$, we have

$$\big((\cdot)^T \otimes id\big)(|\Omega\rangle\langle\Omega|) = \sum_{i,j} |j\rangle\langle i| \otimes |i\rangle\langle j| = \sum_{i,j} |j\rangle|i\rangle\langle i|\langle j|$$

which evaluates to -2 when $\langle i|\langle j| - \langle j|\langle i|$ and $|i\rangle|j\rangle - |j\rangle|i\rangle$ are applied. In order to argue positivity, we recall that the transposition map preserves the eigenvalues of Hermitian operators (and the property of being Hermitian).

Below, we will show different equivalent characterizations of CPTP maps; one of them will in particular show that the CPTP property is not only necessary but also sufficient for a map to be a quantum operation, i.e., to be "physically implementable".

We start off with the following concept.

[3] Formally, this may cause some ambiguity when E depends on several random variables X, Y etc; however, the meaning should always be clear from the context: e.g., ρ_E^y will always stand for $\rho_{E|Y=y}$, and not for $\rho_{E|X=y}$.

Definition 6.7. *Let $\mathcal{H}_A$ and $\mathcal{H}_{A'}$ be Hilbert spaces and $\{|i\rangle\}_{i\in I}$ an orthonormal basis of $\mathcal{H}_A$. Then, the* **Choi-Jamiołkowski isomorphism** *(w.r.t. $\{|i\rangle\}_{i\in I}$) is the map*

$$J : \mathcal{L}\big(\mathcal{L}(A), \mathcal{L}(A')\big) \to \mathcal{L}(A'A)\,,\ \mathfrak{T} \mapsto (\mathfrak{T} \otimes id)(|\Omega\rangle\langle\Omega|)\,,$$

where $|\Omega\rangle := \sum_i |i\rangle|i\rangle$.

At first glance it may be surprising that $J(\mathfrak{T})$, which, in a sense, is obtained by applying $\mathfrak{T}$ to *one* input, still contains all information on $\mathfrak{T}$. But then, writing out $|\Omega\rangle\langle\Omega|$, we see that

$$J(\mathfrak{T}) = \sum_{i,j} \mathfrak{T}(|i\rangle\langle j|) \otimes |i\rangle\langle j|$$

encodes the action of $\mathfrak{T}$ on a basis of $\mathcal{L}(\mathcal{H}_A)$, and thus on all of $\mathcal{L}(\mathcal{H}_A)$.

Theorem 6.6. *For any $\mathfrak{T} \in \mathcal{L}\big(\mathcal{L}(A), \mathcal{L}(A')\big)$ the following statements are equivalent.*

1. *$\mathfrak{T}$ is a CPTP map.*
2. *$J(\mathfrak{T}) \geq 0$ and $\mathrm{tr}_{A'}\big(J(\mathfrak{T})\big) = \mathbb{I}$ (for any/some choice of basis).*
3. *$\exists\, T_1, \ldots, T_d \in \mathcal{L}(\mathcal{H}_A, \mathcal{H}_{A'}) : \sum_\ell T_\ell^\dagger T_\ell = \mathbb{I}$ and $\mathfrak{T}(R) = \sum_\ell T_\ell R T_\ell^\dagger$ for all $R \in \mathcal{L}(A)$.*
4. *There exists a Hilbert space $\mathcal{H}_E$ and an isometry $V \in \mathcal{L}(A, EA')$ such that*

$$\mathfrak{T}(R) = \mathrm{tr}_{AE} \circ V(R)$$

for all $R \in \mathcal{L}(A)$.

The operator $J(\mathfrak{T})$ is called the **Choi-Jamiołkowski representation** of a CPTP map $\mathfrak{T}$, the family $\{T_1, \ldots, T_d\}$ of operators as in 3. is called the **Kraus representation** of $\mathfrak{T}$, and writing $\mathfrak{T}$ as in 4. is referred to as its **Stinespring representation**.

Remark 6.3. One could phrase (and prove) Theorem 6.6 in a more fine-grained manner. For instance, the trace-preserving property of a CPTP map is related to the partial-trace condition of the Choi-Jamiołkowski representation and to $\sum_\ell T_\ell^\dagger T_\ell = \mathbb{I}$ for the Kraus representation, etc. However, Theorem 6.6 is good enough for our purposes in its current form.

Proof. We prove that 1. ⇒ 2. ⇒ 3. ⇒ 4. ⇒ 1.
1. ⇒ 2. (CPTP ⇒ Choi-Jamiołkowski):

$J(\mathfrak{T}) \geq 0$ follows immediately from the CP property, given that $|\Omega\rangle\langle\Omega| \geq 0$. Furthermore, by linearity of the trace and the TP property:

$$\mathrm{tr}_{A'}\big(J(\mathfrak{T})\big) = \sum_{i,j} \mathrm{tr}\big(\mathfrak{T}(|i\rangle\langle j|)\big) \otimes |i\rangle\langle j| = \sum_{i,j} \mathrm{tr}(|i\rangle\langle j|) \otimes |i\rangle\langle j| = \sum_{i,j} \langle j|i\rangle \otimes |i\rangle\langle j| = \sum_i |i\rangle\langle i| = \mathbb{I}\,.$$

2. ⇒ 3. (Choi-Jamiołkowski ⇒ Kraus):

By assumption $J(\mathfrak{T}) \geq 0$, and thus it has a spectral decomposition $J(\mathfrak{T}) = \sum_\ell \lambda_\ell |e_\ell\rangle\langle e_\ell|$ with $\lambda_\ell \geq 0$. Because of the latter, by rescaling the $|e_\ell\rangle$ we can assume the λ_ℓ to be 1, and invoking the notation introduced in Section 0.7, we can thus write $J(\mathfrak{T}) = \sum_\ell |T_\ell\rangle\langle T_\ell|$ for *operators* $T_\ell \in \mathcal{L}(A, A')$. By construction of $J(\mathfrak{T})$ and by properties discussed in Section 0.7, we see that

$$\mathfrak{T}\big(|i\rangle\langle j|\big) = (\mathbb{I} \otimes \langle i|) J(\mathfrak{T}) (\mathbb{I} \otimes |j\rangle) = \sum_\ell (\mathbb{I} \otimes \langle i|) |T_\ell\rangle\langle T_\ell| (\mathbb{I} \otimes |j\rangle) = \sum_\ell T_\ell |i\rangle\langle j| T_\ell^\dagger$$

for any $i, j \in I$. Furthermore, introducing $\mathbb{I} = \sum_k |k\rangle\langle k| \in \mathcal{L}(A')$,

$$\overline{\sum_\ell T_\ell^\dagger T_\ell} = \sum_\ell \bar{T}_\ell^\dagger \bar{T}_\ell = \sum_{k,\ell} T_\ell^T |k\rangle\langle k| \bar{T}_\ell = \sum_{k,\ell} (\langle k| \otimes \mathbb{I}) |T_\ell\rangle\langle T_\ell| (|k\rangle \otimes \mathbb{I})$$
$$= \sum_k (\langle k| \otimes \mathbb{I}) J(\mathfrak{T}) (|k\rangle \otimes \mathbb{I}) = \mathrm{tr}_A\big(J(\mathfrak{T})\big) = \mathbb{I}\,,$$

exploiting that $\sum_k \langle k| \cdot |k\rangle = \mathrm{tr}$.

3. $\Rightarrow$ 4. (Kraus $\Rightarrow$ Stinespring):

We choose $\mathcal{H}_E$ large enough so that we can select an orthonormal set of vectors $\{|1\rangle, \ldots, |d\rangle\}$ in $\mathcal{H}_E$ with index set as for the Krauss operators $T_1, \ldots, T_d$. Consider then the operator V in $\mathcal{L}(A, EA')$ given by

$$V|\varphi\rangle = \sum_\ell |\ell\rangle \otimes T_\ell |\varphi\rangle\,.$$

Since for arbitrary $|\varphi\rangle, |\psi\rangle \in \mathcal{H}_A$, we have that

$$\langle\psi|\, V^\dagger V |\varphi\rangle = \sum_i \langle\psi| T_i^\dagger T_i |\varphi\rangle = \langle\psi|\varphi\rangle\,,$$

we see that V is an isometry, which proves the claim.

4. $\Rightarrow$ 1. (Stinespring $\Rightarrow$ CPTP): This holds trivially, given that any isometry V, when understood as superoperator $R \mapsto V(R) = VRV^\dagger$, is CPTP, and so is the (partial) trace. $\square$