Contents

Chapter 1

Quantum tools and a first protocol

This chapter covers our first cryptogaphic protocol: we will learn how to encrypt quantum states. To prepare our entry into quantum communication and cryptography we first need to learn a little more about quantum information. Before proceeding, make sure you are comfortable with the notions introduced in Chapter 0.

1.1 Probability notation

There are many good textbooks on probability theory available, such as [**kelly1994introduction**, **ross2010first**], and we refer you to any of them for additional background. Here we recall standard notation which we use throughout.

Consider a discrete random variable X taking values in a set $\mathcal{X}$. We often write $|X|$ for the size of the set $\mathcal{X}$ over which X ranges. The distribution of X is specified by a function $P_X(\cdot) : \mathcal{X} \to [0,1]$ such that for any $x \in \mathcal{X}$, $P_X(x)$ denotes the probability that X takes on a specific value $x \in \mathcal{X}$. Recall that for a probability distribution, the normalization condition $\sum_{x\in\mathcal{X}} P_X(x) = 1$ always holds. When the distribution is clear from context, we use the shorthands

$$p_x = p(x) = \Pr(X = x) = P_X(x) \ .$$

Example 1.1.1. *Let $\mathcal{X} = \{1,2,3,4,5,6\}$ correspond to the faces of a six-sided die. If the die is fair, i.e. all sides have equal probability of occurring, then $P_X(x) = 1/6$ for all $x \in \mathcal{X}$. Using our shorthand notation this can also be written as $p_x = p(x) = 1/6$. The size of the range of X is given by $|X| = 6$.* ■

A random variable X ranging over a set $\mathcal{X}$ can be correlated with another random variable Y ranging over $\mathcal{Y}$. This means that they have a joint distribution $P_{XY}(\cdot,\cdot) : \mathcal{X} \times \mathcal{Y} \to [0,1]$ that is not necessarily a product, that is, $P_{XY}(x,y) \neq P_X(x)P_Y(y)$ in general, where P_X (resp. P_Y) is the marginal distribution of X (resp. Y), defined by $P_X(x) = \sum_{y\in\mathcal{Y}} P_{XY}(x,y)$ (and similarly for Y). This leads to the notion of *conditional probabilities* $P_{X|Y}(x|y)$, where $P_{X|Y}(x|y)$ is the probability that X takes on the value x, conditioned on the event that Y takes on the value y. Bayes' rule relates this conditional probability to the joint probabilities.

$$P_{X|Y}(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)} \ , \tag{1.1}$$

whenever $P_Y(y) > 0$ [1]. We use the following shorthands when it is clear which random variable we refer to:

$$p_{x|y} = p(x|y) = \Pr(X = x|Y = y) = P_{X|Y}(x|y). \tag{1.2}$$

Example 1.1.2. *Let* $Y \in \mathcal{Y} = \{\text{“fair”}, \text{“unfair”}\}$ *refer to the choice of either a fair or an unfair die, each chosen with equal probability:* $P_Y(\text{fair}) = 1/2$ *and* $P_Y(\text{unfair}) = 1/2$*. If* X *denotes the fair or unfair die, where the unfair die always rolls a “6” (that is,* $\mathcal{X} = \{1,2,3,4,5,6\}$*, with* $P_X(6) = 1$ *and* $P_X(x) = 0$ *for* $x \neq 6$*), then* $P_{X|Y}(x|\text{fair}) = 1/6$*,* $P_{X|Y}(6|\text{unfair}) = 1$ *and* $P_{X|Y}(x|\text{unfair}) = 0$ *for* $x \neq 6$*.* ■

Exercise 1.1.1. *Compute explicitly the joint probability* $P_{XY}(x,y)$ *for the previous example.* ■

Exercise 1.1.2. *Suppose that we choose between the fair or unfair die with probability* $P_Y(\text{fair}) = P_Y(\text{unfair}) = 1/2$*, but don't reveal which choice was wade. Then, we roll the (fair or unfair) die and reveal the outcome* X*. Suppose that* $X = 3$*. What is the most likely value of* Y*, “fair” or “unfair”? What if* $X = 6$*?* ■

1.2 Density matrices

The quantum generalization of probability distributions, i.e. probability distributions over quantum states, are called *density matrices*. There are two motivations for considering density matrices. Let's start with the question of how to write down the state of one of several qubits. To this end, imagine given two quantum systems A and B. For example, A and B are two qubits such that the state of A and B is a normalized vector $|\psi\rangle_{AB} \in \mathbb{C}^4$. Given this situation, how would you describe the

[1] Note that the distribution over x given y is irrelevant if y cannot occur $P_Y(y) = 0$.

state of qubit A? Note that physically speaking, if we imagine qubits A and B as being in very far-away locations, then intuitively there must be a way to describe the state of A without referring to B at all.

Consider first an easy case. Suppose that the joint state of A and B takes the form

$$|\psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B \ .$$

Then the answer is clear: the state of A is the normalized vector $|\psi\rangle_A$. However, remember from Chapter 0 that there exists quantum states $|\psi\rangle_{AB}$ that are defined as a superposition of tensor product states in a way that makes it non-obvious whether the state can be directly written as a single tensor product. A good example of such a state is the EPR pair

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}\,|0\rangle_A\,|0\rangle_B + \frac{1}{\sqrt{2}}\,|1\rangle_A\,|1\rangle_B\,.$$

As shown in Exercise **??**, it is impossible to express $|\psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B$ for some states $|\psi\rangle_A$ and $|\psi\rangle_B$. In this case, how can we describe the state of A? It seems like we dug ourselves into a mathematical rabbit-hole. Either we find a way to describe the state of A, or there is a problem with our formalism. As we will see, the notion of density matrix will help us save the day.

The second motivation for introducing density matrices arises when we try to model a probabilistic process using our formalism. Suppose for example that we build a device that prepares either a state $|\psi_1\rangle$, with some probability p_1, or a state $|\psi_2\rangle$, with probability p_2. Wouldn't it be nice to have a concise mathematical way to describe the quantum state returned by this device, every time it is executed, without having to resort to words as in the previous sentence?

1.2.1 Introduction to the formalism

We start by giving a different way to represent pure quantum states, as matrices. Recall that a ket $|\psi\rangle$ is a column vector, while the bra $\langle\psi|$ is a row vector. Therefore, $\rho = |\psi\rangle\langle\psi|$ is a rank-1 matrix: it has precisely 1 non-zero eigenvalue (equal to 1) with associated eigenstate $|\psi\rangle$. The matrix ρ is called the *density matrix* of the quantum state.

Example 1.2.1. *For the states $|0\rangle$ and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ we obtain the matrices*

$$|0\rangle\langle 0| = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (1\ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} , \tag{1.3}$$

$$|+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} (1\ 1) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} . \tag{1.4}$$

■

How does writing down states as matrices help us to resolve the questions above? To see how, let us first consider the second motivation given earlier: the need for a formalism that can represent probabilistic combinations of pure quantum states. But before that, let us remember that the only information that we can gain about a quantum state is obtained by performing a measurement. Moreover, if a state $|\psi\rangle$ is measured in a basis that contains the vector $|b\rangle$, then the probability of obtaining the outcome '$|b\rangle$' is given by

$$|\, \langle b|\, \psi\rangle|^2 = \langle b|\, \psi\rangle\langle \psi\, |b\rangle = \langle b|\, \rho\, |b\rangle \ , \tag{1.5}$$

where as earlier $\rho = |\psi\rangle \langle\psi|$ is the matrix representation of $|\psi\rangle$. In words: the probability of obtaining the outcome '$|b\rangle$' is obtained by computing the *overlap* of $|b\rangle$ with ρ.

Moving on, let us consider the case where our preparation device prepares one of two possible states, $|\psi_1\rangle$ or $|\psi_2\rangle$, with equal probability $p_1 = p_2 = 1/2$. We claim that an accurate matrix representation of the state produced by the device can be obtained by taking the linear combination

$$\rho = \frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2| \ . \tag{1.6}$$

More generally, if the device prepares $|\psi_x\rangle$ with probability p_x, the density matrix representation of the resulting state is

$$\rho = \sum_x p_x|\psi_x\rangle\langle\psi_x| \ . \tag{1.7}$$

To verify that this choice of representation is accurate, consider what happens if we measure the state output by the device in a basis that contains the vector $|b\rangle$. If the state is some $|\psi_x\rangle$, then the probability of obtaining the outcome '$|b\rangle$' is

$$q_{b|x} = |\, \langle b|\, \psi_x\rangle|^2 = \langle b|\psi_x\rangle\langle\psi_x|0\rangle \ . \tag{1.8}$$

Since state $|\psi_x\rangle$ is prepared with probability p_x, we expect the overall probability of obtaining the outcome '$|b\rangle$' to be

$$q_b = \sum_x p_x q_{0|x} \ . \tag{1.9}$$

Observe that

$$q_b = \sum_x p_x q_{b|x} = \sum_x p_x\langle b|\psi_x\rangle\langle\psi_x|b\rangle = \langle b| \left(\sum_x p_x|\psi_x\rangle\langle\psi_x|\right) |b\rangle = \langle b|\, \rho\, |b\rangle \ , \tag{1.10}$$

which is precisely the same rule as (1.5). This means that the density matrix representation posited in (1.7) captures the right amount of information about the state of the system so that the distribution of outcomes of any measurement on the state can be recovered using the rule (1.5).

Example 1.2.2. *Suppose more generally that the device prepares a state with density matrix ρ_x with probability p_x. Then density matrix that describes the overall state prepared by the device is given by*

$$\rho = \sum_x p_x \rho_x \ . \tag{1.11}$$

The set of probabilities and density matrices $\mathcal{E} = \{(p_x, \rho_x)\}_x$ is called an ensemble *of states.* ■

Example 1.2.3. *Suppose the source prepares $|0\rangle\langle 0|$ with probability $1/2$, and $|+\rangle\langle +|$ with probability $1/2$. Then the resulting density matrix for the ensemble $\{(1/2, |0\rangle\langle 0|), (1/2, |+\rangle\langle +|)\}$ is given by*

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} . \tag{1.12}$$

■

Beware that superposition is not the same as a mixture! Intuitively, the difference is that a mixture is an inherently classical object: there is a process that prepares one *or* the other state with some probability. In contrast, a state in a superposition is one *and* the other; it is a truly quantum phenomenon. The following example demonstrates the difference between the two.

Example 1.2.4. *Consider the difference between preparing a* mixture *of $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$, or creating a* superposition *over $|0\rangle$ and $|1\rangle$. First consider a source that prepares the states $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$ with probabilities $p_0 = p_1 = 1/2$. Suppose we measure the resulting state*

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \mathbb{I}/2 \tag{1.13}$$

in the Hadamard basis $\{|+\rangle, |-\rangle\}$. Then the probability of each possible outcome is given by

$$q_+ = \langle +|\,\rho\,|+\rangle = \frac{1}{2} \ , \tag{1.14}$$

$$q_- = \langle -|\,\rho\,|-\rangle = \frac{1}{2} \ . \tag{1.15}$$

In contrast, consider now a state that is an equal superposition of $|0\rangle$ *and* $|1\rangle$*, i.e. the state* $|+\rangle$*. Measuring* $|+\rangle$ *in the Hadamard basis results in* $q_+ = 1$ *and* $q_- = 0$*. The probabilities are different, so the two states are different!* ■

Remark 1.2.1. *Note that the same density matrix can be obtained from different ensembles. A simple example is provided by the density matrix*

$$\rho = \frac{\mathbb{I}}{2} \,, \tag{1.16}$$

which is also called the maximally mixed *state. You can verify that*

$$\frac{\mathbb{I}}{2} = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right) = \frac{1}{2}\left(|+\rangle\langle +| + |-\rangle\langle -|\right) \,, \tag{1.17}$$

and many other equivalent decompositions are possible. (The maximally mixed state arises very frequently in cryptography, because it represents a state of complete uncertainty.*)*

1.2.2 A little bit of math

To formally define density matrices and their properties, we recall some important notions from linear algebra. The first term we introduce is that of a *linear operator*, which in our context is a fancy name to designate a matrix. The term highlights the idea that a matrix maps one vector to another, so it can be thought of as an operation performed on vectors, which – since matrix multiplication is linear – is a linear operation.

Definition 1.2.1 (Linear operator)**.** *A* linear operator $L : \mathbb{C}^d \to \mathbb{C}^{d'}$ *can be represented as a* $d' \times d$ *matrix,*

$$L = \begin{pmatrix} L_{11} & L_{12} & \cdots & L_{1d} \\ L_{21} & \ddots & \ddots & L_{2d} \\ \vdots & \ddots & \ddots & \vdots \\ L_{d'1} & L_{d'2} & \cdots & L_{d'd} \end{pmatrix} , \tag{1.18}$$

where each element $L_{ij} \in \mathbb{C}$*. The linear operator sends the vector* $|\psi\rangle = \sum_{j=1}^{d} \alpha_j |j\rangle$ *to the vector*

$$L|\psi\rangle = \sum_{i=1}^{d'} \Big(\sum_{j=1}^{d} L_{ij}\alpha_j \Big) |i\rangle \ .$$

The set of linear operators from $\mathbb{C}^d$ *to* $\mathbb{C}^{d'}$ *is denoted* $\mathcal{L}(\mathbb{C}^d, \mathbb{C}^{d'})$*.*

Definition 1.2.2 (Hermitian matrix M)**.** *A linear operator* $M \in \mathcal{L}(\mathbb{C}^d, \mathbb{C}^d)$ *is* Hermitian *if* $M^\dagger = M$.

The spectral theorem states that any Hermitian operator $M \in \mathcal{L}(\mathbb{C}^d, \mathbb{C}^d)$ can be diagonalized with real eigenvalues. This means that there exists an orthonormal basis $\{|v_j\rangle\}$ of $\mathbb{C}^d$ (the *eigenvectors*) and real numbers λ_j (the *eigenvalues*) such that $M = \sum_j \lambda_j |v_j\rangle\langle v_j|$.

Definition 1.2.3 (Positive semidefinite matrix)**.** *A Hermitian matrix* M *is* positive semidefinite *if all its eigenvalues* $\{\lambda_i\}_i$ *are non-negative. This condition is denoted as* $M \geq 0$.

Exercise 1.2.1. *Show that a matrix* M *is positive semidefinite if and only if* $\langle v| M |v\rangle \geq 0$ *for all unit vectors* $|v\rangle$*. In particular, the diagonal coefficients* $\langle i| M |i\rangle$ *of* M *in any basis are non-negative. Show that this is not a sufficient condition: find an* M *such that the diagonal coefficients of* M *are all positive but* M *itself is not positive semidefinite. Even worse: find an* M *such that* all *coefficients of* M *are non-negative, but* M *is not positive semidefinite.* ■

An important operation on matrices is the *trace*, which is simply the sum of the diagonal elements. It is convenient to note that the trace can also be expressed as follows.

Definition 1.2.4 (Trace of a matrix)**.** *The trace of a matrix* $M \in \mathcal{L}(\mathbb{C}^d, \mathbb{C}^d)$ *is*

$$\mathrm{tr}(M) = \sum_i \langle i| M |i\rangle \,,$$

where $\{|i\rangle\}$ *is* any *orthonormal basis of* $\mathbb{C}^d$.

You should convince yourself that the definition of the trace does not depend on the choice of orthonormal basis! In particular, by choosing the basis of eigenvectors of M, you can verify that for any Hermitian matrix M, $\mathrm{tr}(M)$ is the sum of its eigenvalues (counted with multiplicity).

The following exercise establishes an important property of the trace, which is that it is *cyclic*.

Exercise 1.2.2. *Show that for any matrices* M, N *we have* $\mathrm{tr}(MN) = \mathrm{tr}(NM)$. *We will often use this property to perform manipulations such as*

$$\langle i| M |i\rangle = \mathrm{tr}(\langle i| M |i\rangle) = \mathrm{tr}(M|i\rangle\langle i|). \tag{1.19}$$

(Make sure you can follow all the kets and bras!) It is worth noting that in general, a non-cyclic permutation of the matrices does not preserve the trace. More precisely, for matrices M, N, P, in general

$$\mathrm{tr}(MNP) \neq \mathrm{tr}(NMP). \tag{1.20}$$

■

1.2.3 Density matrices and their properties

Before we take the density matrix ρ as our new definition for a general quantum sate, let us first investigate when an arbitrary matrix ρ can be considered a valid density matrix, that is, a description of a quantum state. It turns out that two properties are necessary and sufficient in order for a density matrix to represent a valid quantum state: the matrix should be *positive semidefinite* and have *trace equal to 1*.

To see why this is true, consider the diagonalized representation of a density matrix ρ as a function of its eigenvalues $\{\lambda_j\}_j$ and corresponding eigenvectors $\{|v_j\rangle\}_j$:

$$\rho = \sum_j \lambda_j |v_j\rangle\langle v_j| \ . \tag{1.21}$$

Imagine that we measure ρ in an orthonormal basis $\{|w_k\rangle\}_k$. Based on (1.5) we know that the probability of obtaining measurement outcome k should be given by

$$q_k = \langle w_k| \, \rho \, |w_k\rangle \ . \tag{1.22}$$

For this to specify a proper distribution, it must be that $q_k \geq 0$ and $\sum_k q_k = 1$. By performing the measurement in the eigenbasis of ρ, $|w_j\rangle = |v_j\rangle$, we obtain the necessary conditions $\lambda_j \geq 0$, that is, ρ is a *positive semidefinite* matrix, and $\mathrm{tr}(\rho) = 1$, since

$$1 = \sum_j q_j = \sum_j \lambda_j \, \mathrm{tr}(|v_j\rangle\langle v_j|) = \mathrm{tr}(\rho) \ . \tag{1.23}$$

This shows that the two conditions are necessary for ρ to lead to well-defined distributions on measurement outcomes when using the rule (1.5). The following exercise asks you to show that the conditions are also sufficient.

Exercise 1.2.3. *Show that for any positive semidefinite matrix ρ with trace 1, and any orthonormal basis $\{|w_k\rangle\}_k$, the numbers $q_k = \langle w_k| \, \rho \, |w_k\rangle$ are real, non-negative, and sum to 1.* ■

We give a formal definition of a density matrix, which is the most general way of representing a quantum state.

Definition 1.2.5 (Density matrix)**.** *A* density matrix *on* $\mathbb{C}^d$ *is a linear operator* $\rho \in \mathcal{L}(\mathbb{C}^d, \mathbb{C}^d)$ *such that* $\rho \geq 0$ *and* $\mathrm{tr}(\rho) = 1$*. If furthermore* ρ *is of rank* 1*, then* ρ *is called a* pure *state. Otherwise it is called* mixed*.*

Note that by definition a pure density matrix is of the form $\rho = \lambda_1 |u_1\rangle\langle u_1|$, where the trace condition implies that necessarily $\lambda_1 = 1$. Thus for the case of pure states, density matrices and the vector representation we got used to before are in one-to-one correspondence.

We also summarize the rule for computing outcome probabilities when measuring a quantum system described by the density matrix ρ.

Definition 1.2.6 (Measuring a density matrix in a basis)**.** *Consider a density matrix* ρ*. Measuring* ρ *in the orthonormal basis* $\{|b_j\rangle\}_j$ *results in outcome* j *with probability*

$$q_j = \langle b_j| \, \rho \, |b_j\rangle \ . \tag{1.24}$$

1.2.4 Bloch representation for one qubit mixed states

In Chapter 0 we saw that single-qubit states have a convenient graphical representation in terms of a vector on the Bloch sphere. In particular, any pure quantum state can be described by a *Bloch vector* $\vec{r} = (\cos\phi \sin\theta, \sin\phi\sin\theta, \cos\theta)$. As it turns out, the representation extends to mixed states as well. Concretely, we can write any single-qubit density matrix as

$$\rho = \frac{1}{2}\left(\mathbb{I} + v_x X + v_z Z + v_y Y\right) \ , \tag{1.25}$$

where X, Y, Z are the Pauli matrices defined in Chapter 0. The fact that such an expansion always exists follows from the fact that the matrices $\mathcal{S} = \{\mathbb{I}, X, Z, Y\}$ form a basis for the space of 2×2 density matrices that correspond to a qubit. You should convince yourself that all these matrices are orthogonal under the Hilbert-Schmidt inner product $\langle A, B\rangle = \mathrm{tr}(A^\dagger B)$. That is,

$$\mathrm{tr}(X^\dagger Y) = \mathrm{tr}(X^\dagger Z) = \mathrm{tr}(X^\dagger \mathbb{I}) = 0 \ , \tag{1.26}$$

and similarly for all other pairs of matrices.

Exercise 1.2.4. *Use the fact that all matrices* $M, N \in \mathcal{S}$ *with* $M \neq N$ *anti-commute, i.e.,* $\{M, N\} = MN + NM = 0$ *to show that* $\mathrm{tr}(MN) = 0$ *whenever* $M \neq N \in \mathcal{S}$*.* ■

Exercise 1.2.5. *Using the orthogonality condition* (1.26)*, show that*

$$|0\rangle\langle 0| = \frac{1}{2}\left(\mathbb{I} + Z\right) \ , \tag{1.27}$$

$$|1\rangle\langle 1| = \frac{1}{2}\left(\mathbb{I} - Z\right) \ , \tag{1.28}$$

■

If ρ is pure, you can verify that the vector $\vec{v} = (v_x, v_y, v_z)$ is precisely the Bloch vector $\vec{r}$ defined in Chapter 0. For pure states $\|\vec{v}\|_2^2 = v_x^2 + v_y^2 + v_y^2 = 1$. For mixed states, however, we can have $\|\vec{v}\|_2^2 \leq 1$. Thus for the case of 2×2 matrices the vector $\vec{v}$ tells us immediately whether the matrix ρ is a valid one qubit quantum state: this is the case if and only if $\|\vec{v}\|_2 \leq 1$.

1.2.5 Combining density matrices

Suppose given two quantum systems A and B, described by density matrices ρ_A and ρ_B respectively. How should their joint state ρ_{AB} be defined? In the previous chapter we saw that two pure quantum states which can be represented by vectors $|v_1\rangle \in \mathbb{C}^{d_1}, |v_2\rangle \in \mathbb{C}^{d_2}$ respectively can be combined by taking their tensor product $|v_1\rangle \otimes |v_2\rangle \in \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$. It turns out that the rule for mixed states is very similar, and a simple extension of the rule for taking tensor products of vectors.

Let us start with the simple case where ρ_A, ρ_B are 2×2-dimensional matrices. Then

$$\rho_A \otimes \rho_B = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \otimes \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} = \begin{pmatrix} m_{11}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} & m_{12}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \\ m_{21}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} & m_{22}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \end{pmatrix} \tag{1.29}$$

$$= \begin{pmatrix} m_{11}n_{11} & m_{11}n_{12} & m_{12}n_{11} & m_{12}n_{12} \\ m_{11}n_{21} & m_{11}n_{22} & m_{12}n_{21} & m_{12}n_{22} \\ m_{21}n_{11} & m_{21}n_{12} & m_{22}n_{11} & m_{22}n_{12} \\ m_{21}n_{21} & m_{21}n_{22} & m_{22}n_{21} & m_{22}n_{22} \end{pmatrix} . \tag{1.30}$$

For example, if we have two density matrices $\rho_A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\rho_B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, then

$$\rho_{AB} = \rho_A \otimes \rho_B = \begin{pmatrix} 1 \cdot \rho_B & 0 \cdot \rho_B \\ 0 \cdot \rho_B & 0 \cdot \rho_B \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} . \tag{1.31}$$

This definition easily extends to larger matrices as follows:

Definition 1.2.7 (Tensor product)**.** *Consider any $d' \times d$ matrix ρ_A and $k' \times k$ matrix ρ_B,*

$$\rho_A = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1d} \\ m_{21} & \ddots & \ddots & m_{2d} \\ \vdots & \ddots & \ddots & \vdots \\ m_{d'1} & m_{d'2} & \cdots & m_{d'd} \end{pmatrix}, \qquad \rho_B = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1k} \\ n_{21} & \ddots & \ddots & n_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ n_{k'1} & n_{k'2} & \cdots & n_{k'k} \end{pmatrix}. \tag{1.32}$$

Their tensor product is given by the $d'k' \times dk$ matrix

$$\rho_{AB} = \rho_A \otimes \rho_B = \begin{pmatrix} m_{11}\rho_B & m_{12}\rho_B & \cdots & m_{1d}\rho_B \\ m_{21}\rho_B & \ddots & \ddots & m_{2d}\rho_B \\ \vdots & \ddots & \ddots & \vdots \\ m_{d'1}\rho_B & m_{d'2}\rho_B & \cdots & m_{d'd}\rho_B \end{pmatrix}. \tag{1.33}$$

As a word of caution, we note that the tensor product, as the usual matrix product, is non-commutative.

Example 1.2.5. *Consider the density matrices $\rho_A = \frac{1}{4}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$ and $\rho_B = \frac{1}{2}\begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}$. Then*

$$\rho_A \otimes \rho_B = \frac{1}{8}\begin{pmatrix} 1 & -i & 1 & -i & 0 & 0 \\ i & 1 & i & 1 & 0 & 0 \\ 1 & -i & 2 & -2i & 1 & -i \\ i & 1 & 2i & 2 & i & 1 \\ 0 & 0 & 1 & -i & 1 & -i \\ 0 & 0 & i & 1 & i & 1 \end{pmatrix}, \tag{1.34}$$

and

$$\rho_B \otimes \rho_A = \frac{1}{8}\begin{pmatrix} 1 & 1 & 0 & -i & -i & 0 \\ 1 & 2 & 1 & -i & -2i & -i \\ 0 & 1 & 1 & 0 & -i & -i \\ i & i & 0 & 1 & 1 & 0 \\ i & 2i & i & 1 & 2 & 1 \\ 0 & i & i & 0 & 1 & 1 \end{pmatrix} \neq \rho_A \otimes \rho_B. \tag{1.35}$$

■

1.2.6 Classical-quantum states

In quantum cryptography we frequently find ourselves in a situation in which the honest parties have some classical information X about which an adversary — such as an eavesdropper Eve — may hold quantum information Q. So the quantum state Q is correlated with the classical information. Since classical information is a special case of quantum information, the joint state of both X and Q can be represented by a density matrix ρ_{XQ}. How does such a density matrix look like?

Classical states

As a first step, let us pause to think about what it means for X to contain "classical information". In full generality, classical information can be modeled by a probability distribution over strings of bits x: x denotes the information and p_x the probability that this is the information contained in X; this is exactly analogous to the notion of a random variable in classical probability theory.

Suppose then that we are given a probability distribution over symbols from the alphabet $\mathcal{X} = \{0, \dots, d-1\}$, and let p_x denote the probability of symbol x. Identifying each possible value in $\mathcal{X}$ with an element of the standard basis $\{|0\rangle, \dots, |d-1\rangle\}$ we can describe a system that is initialized in state $|x\rangle$ with probability p_x using the density matrix

$$\rho = \sum_{x=0}^{d-1} p_x |x\rangle\langle x| \ . \tag{1.36}$$

You may want to check that measuring ρ in the standard basis results in outcome "x" with probability precisely p_x: in this sense, ρ is an accurate representation of system X as described above.

Note that ρ is a matrix which has the probabilities p_x on the diagonal and is has all other entries zero. As such, ρ is just another way to represent the distribution p_x: instead of a sequence of numbers, or a vector, we wrote the numbers on the diagonal of a matrix.

Definition 1.2.8 (Classical state)**.** *Let $\{|x\rangle\}_{x=0}^{d-1}$ denote the standard basis for $\mathbb{C}^d$. A system X is said to be in a classical state, or* c-state*, if its density matrix ρ_X is diagonal in the standard basis, i.e. ρ_X has the form*

$$\rho = \sum_{x=0}^{d-1} p_x |x\rangle\langle x| \ , \tag{1.37}$$

where $\{p_x\}_{x=0}^{d-1}$ is a probability distribution.

Thus from now on we equate "classical state" or "classical density matrix" with "diagonal in the standard basis". The choice of the standard basis is arbitrary, as from a mathematical point of view all orthonormal bases are equivalent. Nevertheless, it is an important convention and serves as a point of connection between classical and quantum world.

Classical-quantum states

In quantum cryptography we often encounter states which are partially classical and partially quantum. Here is an example of such a state. Suppose that with probability $1/2$ system X is in the classical state $|0\rangle$ and Q is in the mixed state $\frac{\mathbb{I}}{2}$, and with probability $1/2$ system X is in the classical state $|1\rangle$ and Q is in the pure state $|+\rangle$. How do we write down the density matrix of XQ? Following the description, we obtain the formula

$$\rho_{XQ} = \frac{1}{2}|0\rangle\langle 0|_X \otimes \frac{\mathbb{I}_Q}{2} + \frac{1}{2}|1\rangle\langle 1|_X \otimes |+\rangle\langle +|_Q \,. \tag{1.38}$$

As an important exercise to verify that you are following, make sure you are able to confirm that ρ_{XQ} is a valid density matrix (remember the two conditions that need to be checked?). This kind of density matrix is called a *classical-quantum state*, or cq-state for short. The reason is that the X part of the state is always classical. More generally, we give the following definition.

Definition 1.2.9. *A classical-quantum state, or simply* cq-state*, is a state of two subsystems, X and A, such that its density matrix has the form*

$$\rho_{XQ} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_x^Q \,, \tag{1.39}$$

where (p_x) is a probability distribution and for every x, $|x\rangle$ designates the standard basis state on X and ρ_x^Q is an arbitrary density matrix on Q.

In applications to cryptography x will often represent some (partially secret) classical string that Alice creates during a quantum protocol, and ρ_x^Q some quantum information that an eavesdropper may have gathered during the protocol, and which may be correlated with the string x. It is an established custom in the quantum information literature to reserve the letters X, Y, Z to denote classical registers, and use the other letters for quantum information. (More letters for quantum!)

1.3 General measurements

So far we have described how to measure a quantum state in a given basis. Quantum mechanics allows a much more refined notion of measurement, which plays an important role both in quantum information theory and cryptography. Indeed, in quantum information theory certain tasks, such as the task of discriminating between multiple states, can be solved more efficiently using these generalized measurements. Moreover, taking an adversarial viewpoint, in quantum cryptography it is essential to prove security for the most general kind of attack, including all measurements that an attacker could possibly make!

1.3.1 POVMs

If we are only interested in computing the probabilities of measurement outcomes - but do not require a complete specification of what happens to the quantum state once the measurement has been performed - then the most general kind of measurement that is allowed in quantum mechanics can be described by a positive operator-valued measure, or POVM for short.

Definition 1.3.1 (POVM). *A POVM on $\mathbb{C}^d$ is a set of positive semidefinite operators $\{M_x\}_{x\in\mathcal{X}}$ such that*

$$\sum_x M_x = \mathbb{I}_{\mathbb{C}^d} \ . \tag{1.40}$$

The subscript x is used as a label for the measurement outcome. The probability p_x of observing outcome x can be expressed using the Born rule *as*

$$p_x = \mathrm{tr}(M_x\rho) \ . \tag{1.41}$$

Example 1.3.1. *Recall that when measuring a state $|\psi\rangle = \sum_x \alpha_x |x\rangle$ in a basis such as $\{|x\rangle\}_x$, the probability of observing outcome x is given by $|\alpha_x|^2$. Let's see how this rule can be recovered as a special case of the POVM formalism we just introduced. For each x let $M_x = |x\rangle\langle x|$, so that M_x is positive semidefinite (it fact, it is a projector, i.e. $M^2 = M$) and $\sum_x M_x = \mathbb{I}$ (this can be verified by using that $\{|x\rangle\}$ is a basis), as required. We can use the Born rule to compute*

$$\begin{aligned} p_x &= \mathrm{tr}(M_x\rho) \\ &= \mathrm{tr}(|x\rangle\langle x|\rho) \\ &= \langle x|\,\rho\,|x\rangle \\ &= \sum_{x',x''} \alpha_{x'}\alpha^*_{x''}\,\langle x|\,x'\rangle\langle x''\,|x\rangle \\ &= |\alpha_x|^2 \ . \end{aligned}$$

■

Example 1.3.2. *Consider a distribution (p_x) and the associated classical mixture $\rho = \sum_x p_x |x\rangle\langle x|$. If we measure ρ in the standard basis, with associated POVM $M_x = |x\rangle\langle x|$ as in the Example 1.3.1, we obtain outcome x with probability*

$$\mathrm{tr}(|x\rangle\langle x|\rho) = \langle x|\,\rho\,|x\rangle = p_x, \tag{1.42}$$

as expected: ρ indeed captures the classical distribution given by the probabilities p_x. ■

Beyond the calculation of outcome probabilities, it is often important to know what happens to a quantum state after a generalized measurement has been performed. For the case of measuring in a basis, we already know the answer: the state directly collapses to the basis element associated with the outcome of the measurement that is obtained.

In the case of a POVM it turns out that the information given by the operators $\{M_x\}$ is not sufficient to fully determine the post-measurement state. The reason for this is because such a measurement may not fully collapse the state, meaning that the post-measurement state may not be pure (this corresponds to the case where M_x has rank more than 1). Intuitively, if the measurement operator M_x does not have rank 1 there is some freedom in choosing exactly where it lies without affecting the outcome probabilities.

1.3.2 Generalized measurements

In order to specify post-measurement states we need to give a *Kraus operator representation* of the POVM.

Definition 1.3.2 (Kraus operators)**.** *Let $M = \{M_x\}$ be a given POVM on $\mathbb{C}^d$. A* Kraus operator representation *of M is a set of linear operators $A_x \in \mathcal{L}(\mathbb{C}^d, \mathbb{C}^{d'})$ such that $M_x = A_x^\dagger A_x$ for all x.*

Note that a Kraus decomposition of any POVM always exists by setting $A_x = \sqrt{M_x}$, the positive square root of M_x. (For any positive semidefinite matrix N, if $N = \sum_i \lambda_i |v_i\rangle\langle v_i|$ is the spectral decomposition of N, then $\sqrt{N} = \sum_i \sqrt{\lambda_i}|v_i\rangle\langle v_i|$.) In particular, if $M_x = |u_x\rangle\langle u_x|$ is a projector then $\sqrt{M_x} = M_x$ and we can take $A_x = M_x$. But for any unitary U_x on $\mathbb{C}^d$, $A'_x = U_x\sqrt{M_x}$ is also a valid decomposition. Hence, there is no unique Kraus representation for a given POVM.

The most general form to write down a quantum measurement is given by the full set of Kraus operators $\{A_x\}_x$. From these, we can easily find the POVM operators as $M_x = A_x^\dagger A_x$. But knowledge of the Kraus operators allows us to do more: it allows to compute post-measurement states.

Definition 1.3.3 (Post-measurement state)**.** *Let ρ be a density matrix and $M = \{M_x\}$ a POVM with Kraus decomposition given by operators $\{A_x\}$. Suppose the measurement is performed, and the outcome x is obtained. Then the state of the system after the measurement, conditioned on the outcome x, is*

$$\rho_{|x} = \frac{A_x \rho A_x^\dagger}{\mathrm{tr}(A_x^\dagger A_x \rho)} \ .$$

You may want to convince yourself that when measuring a pure state $|\psi\rangle$ in an arbitrary orthonormal basis, with POVM elements $M_x = |x\rangle\langle x|$ and Kraus decomposition $A_x = \sqrt{M_x} = |x\rangle\langle x|$ (the second equality follows since all eigenvalues of a projection are $0, 1$, and in particular their square root is equal to themselves), the post-measurement state as defined above is precisely the basis state associated to the measurement outcome. Note that since a POVM does not in general have a unique decomposition into Kraus operators, specifying POVM operators alone is insufficient to determine the post-measurement state. Nevertheless, talking about a POVM is extremely useful if we only care about measurement probabilities, since the matrices M_x have a slightly simpler form. In particular, as we will note later , we can easily optimize over them using a semidefinite program (SDP).

An important class of generalized measurements is given by the case where the M_x are projectors onto orthogonal subspaces (not necessarily of rank 1).

Definition 1.3.4. *A* projective measurement, *also called a* von Neumann measurement, *is given by a set of orthogonal projectors $M_x = \Pi_x$ such that $\sum_x \Pi_x = \mathbb{I}$. For such a measurement, unless otherwise specified we will always use the default Kraus decomposition $A_x = \sqrt{M_x} = \Pi_x$. The probability q_x of observing measurement outcome x can be expressed as*

$$q_x = \mathrm{tr}(\Pi_x \rho),$$

and the post-measurement states are

$$\rho_{|x} = \frac{\Pi_x \rho \Pi_x}{\mathrm{tr}(\Pi_x \rho)} \ .$$

Example 1.3.3. *Suppose given a two-qubit state ρ, such that we would like to measure the parity (in the standard basis) of the two qubits. A first way to do this would be to measure ρ in the standard basis, obtain two bits, and take their parity. In this case the probability of obtaining the outcome "even" would be*

$$q_{\text{even}} = \langle 00| \rho |00\rangle + \langle 11| \rho |11\rangle \,,$$

and the post-measurement state would be the mixture of the two post-measurement states associated with outcomes $(0,0)$ *and* $(1,1)$*, so*

$$\rho_{|\text{even}} = \big(\langle 00| \, \rho \, |00\rangle \big) \, |00\rangle \langle 00| + \big(\langle 11| \, \rho \, |11\rangle \big) \, |11\rangle \langle 11| \, .$$

Now suppose that we attempt to measure the parity using a generalized measurement which directly projects onto the relevant subspaces, without measuring the qubits individually. That is, consider the projective measurement $\Pi_{\text{even}} = |00\rangle\langle 00| + |11\rangle\langle 11|$ *and* $\Pi_{\text{odd}} = \mathbb{I} - \Pi_{\text{even}} = |01\rangle\langle 01| + |10\rangle\langle 10|$*. With this measurement the probability of obtaining the outcome "even" is*

$$q'_{\text{even}} = \text{tr}(\Pi_{\text{even}}\rho) = \langle 00| \, \rho \, |00\rangle + \langle 11| \, \rho \, |11\rangle \ , \tag{1.43}$$

as before. However, the post-measurement state is now

$$\rho'_{|\text{even}} = \Pi_{\text{even}}\rho\Pi_{\text{even}} \ . \tag{1.44}$$

To see the difference, consider the state $\rho = |\text{EPR}\rangle \langle \text{EPR}|$ *where* $|\text{EPR}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$*. Then clearly the parity measurement should report the outcome "even" with probability* 1*, and you can check that this is the case for both measurements. However, the post-measurement states are different. In the first case,*

$$\rho_{|\text{even}} = \frac{1}{2} \, |00\rangle \langle 00| + \frac{1}{2} \, |11\rangle \langle 11| \, ,$$

while in the second case,

$$\rho'_{|\text{even}} = |\text{EPR}\rangle \langle \text{EPR}|$$

is unchanged! This is one of the key advantages of using generalized measurements, as opposed to basis measurements: they allow to compute certain simple quantities on multi-qubit states (such as the parity) without fully "destroying" the state. ■

Exercise 1.3.1. *Use a projective measurement to measure the parity, in the Hadamard basis, of the state* $|00\rangle\langle 00|$*. Compute the probabilities of obtaining measurement outcomes "even" and "odd", and the resulting post-measurement states. What would the post-measurement states have been if you had first measured the qubits individually in the Hadamard basis, and then taken the parity?* ■

1.4 The partial trace

Going back to our initial motivation for introducing density matrices, let's now give an answer to the following question: given a multi-qubit state, how do we write down the "partial state" associated to a subset of the qubits? More generally, suppose ρ_{AB} is a density matrix on a tensor product space $\mathbb{C}^{d_A}_A \otimes \mathbb{C}^{d_B}_B$, but suppose Alice holds the part of ρ corresponding to system A and Bob holds the part corresponding to system B. How do we describe the state ρ_A of Alice's system?

1.4.1 An operational viewpoint

The operation that takes us from ρ_{AB} to ρ_A is called the *partial trace*. It can be specified in purely mathematical terms, and we do so in the following section. However, before that, let's try to think about the problem from an operational point of view. First, consider an easy case: if $\rho_{AB} = \rho_A \otimes \rho_B$, where ρ_A and ρ_B are both density matrices, then clearly Alice's system is defined by ρ_A. In this case, we'd say that the partial trace of ρ_{AB}, when "tracing out" system B, is the density matrix ρ_A.

A slightly more complicated case is when

$$\rho_{AB} = \sum_i p_i \rho_i^A \otimes \rho_i^B \tag{1.45}$$

is a mixture of tensor products (we will later see this is called a "separable state"). Using the interpretation that this represents a state that is in state $\rho_i^A \otimes \rho_i^B$ with probability p_i, it would certainly be natural to claim that Alice's share of the state is ρ_i^A with probability p_i, i.e. the partial trace of ρ_{AB}, when tracing out system B, is now $\rho_A = \sum_i p_i \rho_i^A$.

How about a general ρ? Remember from Exercise **??** that there exists such ρ that do not have a decomposition of the form (1.45). The idea is to "force" such a decomposition by performing the following little thought experiment. Let's *imagine* that Bob performs a complete basis measurement on his system, using an arbitrary basis $\{|u_x\rangle\}$. Let's introduce a POVM on the joint system of Alice and Bob that models this measurement: since Alice does nothing, we can set $M_x = \mathbb{I}_A \otimes |u_x\rangle\langle u_x|_B$, which you can check indeed defines a valid POVM. Moreover, this is a projective measurement, so we can take the Kraus operators $A_x = \sqrt{M_x} = M_x$. By definition the post-measurement states are given by

$$\rho^{AB}_{|x} = \frac{M_x \rho_{AB} M_x}{\mathrm{tr}\left(M_x \rho_{AB}\right)} = \frac{\big((\mathbb{I}_A \otimes \langle u_x|)\rho_{AB}(\mathbb{I}_A \otimes |u_x\rangle)\big)_A \otimes |u_x\rangle\langle u_x|_B}{\mathrm{tr}\left((\mathbb{I}_A \otimes |u_x\rangle\langle u_x|_B)\rho_{AB}\right)}.$$

Notice how we wrote the state, as a tensor product of a state on A and one on B. Make sure you understand the notation in this formula, and that it specifies a well-defined state.

Now the key step is to realize that, whatever the state of Alice's system A is, it shouldn't depend on any operation that Bob performs on B. After all, it may be that

A is here on earth, and B on Mars. Since quantum mechanics does not allow faster than light communication, as long as the two of them remain perfectly isolated, meaning that Alice doesn't get to learn the measurement that Bob performs or its outcome, then her state should remain unchanged. We can thus describe it as follows: "With probability $q_x = \mathrm{Tr}(M_x \rho_{AB})$, Alice's state is the A part of $\rho^{AB}_{|x}$", i.e.

$$\rho_A = \sum_x q_x \frac{\big((\mathbb{I} \otimes \langle u_x|)\rho_{AB}(\mathbb{I} \otimes |u_x\rangle)\big)_A}{\mathrm{Tr}\,\big((\mathbb{I} \otimes |x\rangle\langle x|)\rho_{AB}\big)} = \sum_x (\mathbb{I} \otimes \langle u_x|)\rho_{AB}(\mathbb{I} \otimes |u_x\rangle). \quad (1.46)$$

Although we derived the above expression for Alice's state using sensible arguments, there is something you should be worried about: doesn't it depend on the choice of basis $\{|u_x\rangle\}$ we made for Bob's measurement? Of course, it should not, as our entire argument is based on the idea that Alice's reduced state should not depend on any operation performed by Bob. (We emphasize that this is only the case as long as Alice doesn't learn the measurement outcome! If we fix a particular outcome x then it's a completely different story; beware of this subtlety, that will repeatedly come up throughout the book.)

Exercise 1.4.1. *Verify that the state ρ_A defined in Eq.(1.46) does not depend on the choice of basis $\{|u_x\rangle\}$. [Hint: first argue that if two density matrices ρ, σ satisfy $\langle\phi|\,\rho\,|\phi\rangle = \langle\phi|\,\sigma\,|\phi\rangle$ for all unit vectors $|\phi\rangle$ then $\rho = \sigma$. Then compute $\langle\phi|\,\rho_A\,|\phi\rangle$, and use the POVM condition $\sum_x M_x = \mathbb{I}$ to check that you can get an expression independent of the $\{|u_x\rangle\}$. Conclude that ρ_A itself does not depend on $\{|u_x\rangle\}$.]* ■

Example 1.4.1. *Consider the example of the EPR pair*

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \quad (1.47)$$

Writing this as a density operator we have

$$\rho_{AB} = |\mathrm{EPR}\rangle\langle\mathrm{EPR}|_{AB} = \frac{1}{2}\left(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\right) . \quad (1.48)$$

Let's measure system B in the standard basis: taking A into account we consider the POVM $M_0 = \mathbb{I}_A \otimes |0\rangle\,\langle 0|_B$ and $M_1 = \mathbb{I}_A \otimes |1\rangle\langle 1|_B$. We can then compute

$$\begin{aligned} q_0 &= \mathrm{Tr}(M_0\rho) \\ &= \frac{1}{2}\,\mathrm{Tr}\,\big((\mathbb{I} \otimes |0\rangle\langle 0|)(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|)\big) \\ &= \frac{1}{2}\big(1 + 0 + 0 + 0\big) = \frac{1}{2}, \end{aligned}$$

and similarly $q_1 = 1/2$. The post-measurement stated on A is then

$$\rho^A_{|0} = \frac{1}{2}(\mathbb{I}\otimes\langle 0|)\rho_{AB}(\mathbb{I}\otimes|0\rangle) + \frac{1}{2}(\mathbb{I}\otimes\langle 1|)\rho_{AB}(\mathbb{I}\otimes|1\rangle) = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|.$$

Exercise: do the same calculation using a measurement in the Hadamard basis on B, and check that you get the same result! ■

1.4.2 A mathematical definition

Armed with our "operational" definition of what the partial trace *should* be, we can now give the precise, mathematical definition of the partial trace operation.

Definition 1.4.1 (Partial Trace)**.** *Consider a general state*

$$\rho_{AB} = \sum_{ijk\ell} \gamma^{k\ell}_{ij} |i\rangle\langle j|_A \otimes |k\rangle\langle \ell|_B, \tag{1.49}$$

where $|i\rangle_A, |j\rangle_A$ and $|k\rangle_B, |\ell\rangle_B$ run over orthonormal bases of A and B respectively. Then the partial trace over B is defined as

$$\rho_A = \operatorname{tr}_B(\rho_{AB}) = \sum_{ijk\ell} \gamma^{k\ell}_{ij} |i\rangle\langle j| \otimes \operatorname{tr}(|k\rangle\langle \ell|) = \sum_{ij}\left(\sum_k \gamma^{kk}_{ij}\right)|i\rangle\langle j| \,. \tag{1.50}$$

Similarly, the partial trace over A is

$$\rho_B = \operatorname{tr}_A(\rho_{AB}) = \sum_{ijk\ell} \gamma^{k\ell}_{ij} \operatorname{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle \ell| = \sum_{k\ell}\left(\sum_j \gamma^{k\ell}_{jj}\right)|k\rangle\langle \ell| \,. \tag{1.51}$$

The states ρ_A, ρ_B are referred to as reduced states.

Example 1.4.2. *Let's consider again the example of the EPR pair*

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle),$$

with associated density matrix $\rho_{AB} = |\mathrm{EPR}\rangle\langle\mathrm{EPR}|_{AB}$. Using the definition we can compute

$$\begin{aligned}\operatorname{tr}_B(\rho_{AB}) = \frac{1}{2}\big(&|0\rangle\langle 0|\otimes\operatorname{tr}(|0\rangle\langle 0|) + |0\rangle\langle 1|\otimes\operatorname{tr}(|0\rangle\langle 1|)\\ &+ |1\rangle\langle 0|\otimes\operatorname{tr}(|1\rangle\langle 0|) + |1\rangle\langle 1|\otimes\operatorname{tr}(|1\rangle\langle 1|)\big)\,.\end{aligned} \tag{1.52}$$

Since the trace is cyclic, $\mathrm{tr}(|0\rangle\langle 1|) = \langle 1|0\rangle = 0$*, similarly* $\mathrm{tr}(|1\rangle\langle 0|) = 0$*, but* $\mathrm{tr}(|0\rangle\langle 0|) = \mathrm{tr}(|1\rangle\langle 1|) = 1$ *and hence*

$$\mathrm{tr}_B(\rho_{AB}) = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right) = \frac{\mathbb{I}}{2}\,. \tag{1.53}$$

Convince yourself that when we take the partial trace operation over A*, and hence look at the state of just Bob's qubit we have*

$$\mathrm{tr}_A(\rho_{AB}) = \frac{\mathbb{I}}{2}\,. \tag{1.54}$$

■

Exercise 1.4.2. *If* $\rho_{AB} = |\Phi\rangle\langle\Phi|$ *is the singlet* $|\Phi\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$*, compute* ρ_A *and* ρ_B*.* ■

Example 1.4.3. *The notion of partial trace allows us to verify that performing a unitary operation on* A *has no effect on the state of* B*, i.e., it does not change* ρ_B*.*

$$(U_A \otimes \mathbb{I}_B)\rho_{AB}(U_A \otimes \mathbb{I}_B)^\dagger = \sum_{ijk\ell} \gamma_{ij}^{k\ell} U_A|i\rangle\langle j|_A U_A^\dagger \otimes |k\rangle\langle\ell|_B. \tag{1.55}$$

Computing again the partial trace we have

$$\begin{aligned}
\mathrm{tr}_A(U_A \otimes \mathbb{I}_B \rho_{AB} U_A^\dagger \otimes \mathbb{I}_B) &= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(U_A|i\rangle\langle j|U_A^\dagger) \otimes |k\rangle\langle\ell| &(1.56)\\
&= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(|i\rangle\langle j|U_A^\dagger U_A) \otimes |k\rangle\langle\ell| &(1.57)\\
&= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle\ell| &(1.58)\\
&= \sum_{k\ell}\left(\sum_j \gamma_{jj}^{k\ell}\right)|k\rangle\langle\ell| = \rho_B\,. &(1.59)
\end{aligned}$$

Can you convince yourself that performing a measurement on A *also has no effect on* B*?* ■

1.5 Secure message transmission

With all the math behind us we are ready to turn to our first serious cryptographic tasks — in fact, the most serious task of all: the secure transmission of messages. To set things up, imagine our favorite protagonists, Alice and Bob. Alice and Bob would like to exchange classical messages (i.e. they just want to chat!), but they are worried that there might be an adversary, also called the eavesdropper Eve, listening on the line. Alice and Bob trust that they each have full control over their own secure laboratory (i.e. their bedroom), that Eve cannot peek into. However, Eve has access to the communication channel connecting Alice and Bob (i.e. she can intercept any cell phone conversation).

In order for Alice and Bob to be able to exchange messages secretely, we need to beak the symmetry somehow: if everyone has access to exactly the same information, then anything Bob can receive, Eve can receive too. To break this symmetry we will make the assumption that Alice and Bob are in possession of a *secret key* that they use to encode their messages. The assumption is that the key is known to both Alice and Bob, but is private to them: Eve has no information at all about it. For this reason we call cryptosystems such as the one we're about to discuss *private-key* cryptosystems. For the time being, we won't justify our assumption about the key: this is just some secret Alice and Bob have in common, a secret they may have agreed on a long time in the past, when they were in the same place and could whisper to each other's ears. Later in the book, we will see how quantum information can also be used to establish such a secret when Alice and Bob are physically separated.

1.5.1 Shannon's secrecy condition and the need for large keys

Write k for Alice and Bob's shared key. To express that Eve has no information whatsoever about the key, we assume that from her point of view every possible key has the same a priori probability: for every k in the key space K, it holds that $p(k) = 1/|K|$. The mathematical framework for the description of secret communication schemes was first developed by Shannon in the 1940s [**shannon1949communication**], well before quantum information made its apparition. According to Shannon's formalism, an encryption scheme consists of two functions. The first is the *encryption function* $\mathrm{Enc}(k, m) = e$, that takes the key k and the message m and maps it to some encrypted message e. The original message m is often called the *plaintext*, and e the *ciphertext*. The second function is the *decryption function* $\mathrm{Dec}(k, e) = m$, that takes the key k and the ciphertext e back to the plaintext.

Definition 1.5.1. *An encryption scheme* $(\mathrm{Enc}, \mathrm{Dec})$ *is called* correct *if for every key k and every plaintext m,* $\mathrm{Dec}(k, \mathrm{Enc}(k, m)) = m$*. It is called* secure *if for any distribution $p(\cdot)$ over the space $\mathcal{M}$ of plaintexts the following two distributions on plaintexts are identical:*

1. *Generate a random plaintext $m \in \mathcal{M}$ with probability $p(m)$.*

2. *Select an arbitrary ciphertext e. Generate a uniformly random key $k \in K$. Generate a random plaintext $m \in \mathcal{M}$ with probability $p(m|\mathrm{Enc}(k,m) = e)$.*

In other words we call an encryption scheme secure whenever an eavesdropper Eve ignorant of the key does not gain any additional information about a plaintext message m when given access to its encryption e: the probability $p(m)$ of the message m is the same a priori (as anyone could guess) as it is from the point of view of Eve, who has obtained e. This is a very strong notion of security: absolutely no information is gained by having access to e! It may even seem impossible to realize: if e has "no information" about m, then how can e be decrypted to recover m? As we will soon see, there is no contradiction: it is possible that e has no information at all about m *from the point of view of an Eavesdropper who does not know the secret key k*, yet e still has full information about m from the point of view of a honest party who does know the secret key. This is a very subtle point: make sure you fully understand the distinction.

Note that it would be easy to come up with an encryption scheme which is "just" secret: Alice simply sends a randomly chosen e to Bob. This is why the correctness requirement is also made explicit in the definition. The art of encryption is to design schemes that are both correct *and* secure.

In our presentation we assumed that Alice and Bob share a secret key k, and we informally argued that such a key was needed to "break the symmetry" between Bob and the eavesdropper Eve. Is this argument watertight — is a key really needed? As it turns out, not only it is needed but in fact the number of possible keys needs to be as large as the number of possible messages that Alice may wish to send. The following lemma, due to Shannon, proves this.

Lemma 1. *An encryption scheme* $(\mathrm{Enc}, \mathrm{Dec})$ *can only be* secure *and* correct *if the number of possible keys $|K|$ is at least as large as the number of possible messages $|M|$, that is, $|K| \geq |M|$.*

Proof. Suppose for contradiction that there exists a correct scheme using fewer keys, i.e., $|K| < |M|$. We will show that such a scheme cannot be secure. Let p be the uniform distribution over M. Consider an eavesdropper who has intercepted the ciphertext e. She can compute

$$\mathcal{S} = \{\hat{m} \mid \exists k, \hat{m} = \mathrm{Dec}(k,e)\} \ , \tag{1.60}$$

that is, the set of all messages $\hat{m}$ for which there exists a key k that could have resulted in the observed ciphertext e. Note that the size $|\mathcal{S}|$ of this set is $|\mathcal{S}| \leq |K|$, since for each possible key k we get at most one message $\hat{m}$. Since $|K| < |M|$, we thus have $|\mathcal{S}| < |M|$. This means that there exists at least one message m such that $m \notin \mathcal{S}$, and hence $p(m|e) = 0$. However, by definition $p(m) = 1/|M|$. This gives a direct contradiction with the definition of security given in Definition 1.5.1. □

Can the bound given in the lemma be achieved: does there exist an encryption scheme that is both correct *and* secure, and which uses precisely the minimal number of keys $|K| = |M|$? The answer is yes! We construct such a scheme in the next section.

1.5.2 The (quantum) one-time pad

The *one-time pad* is arguably the simplest, yet also the most secure, encryption scheme known. We start with the "classical" version, that allows encryption of classical messages.

The classical one-time pad

Imagine that Alice (the sender) wants to send a secret message m to Bob (the receiver). For simplicity, we take the message space M to be the set of all n-bit strings: $M = \{0,1\}^n$. Let us furthermore assume that Alice and Bob already share a key $k \in \{0,1\}^n$ which is just as long as the message, and is uniformly random from the point of view of any adversary. In the following definition, we use the notation $a \oplus b$ for the bitwise XOR, or equivalently addition modulo 2: for $a, b \in \{0,1\}$, $a \oplus b = a + b \bmod 2$.

Protocol 1. *The classical* one-time *pad is an encryption scheme in which the encryption of a message $m \in \{0,1\}^n$ using the key $k \in \{0,1\}^n$ is given by*

$$\mathrm{Enc}(k,m) = m \oplus k = (m_1 \oplus k_1, m_2 \oplus k_2, \ldots, m_n \oplus k_n) = (e_1, \ldots, e_n) = e \, . \tag{1.61}$$

The decryption is given by

$$\mathrm{Dec}(k,e) = e \oplus k = (e_1 \oplus k_1, e_2 \oplus k_2, \ldots, e_n \oplus k_n). \tag{1.62}$$

Note that since for any $j \in \{1, \ldots, n\}$, $m_j \oplus k_j \oplus k_j = m_j$, the scheme is correct. Is it secure?

To see that it satisfies Shannon's definition, consider any distribution p on M. For a uniformly random choice of key k and a fixed message m, the associated ciphertext $e = \mathrm{Enc}(k,m)$ is uniformly distributed over all n-bit strings: for any e,

$$p(\mathrm{Enc}(k,m) = e|m) = p(m \oplus k = e|m) = p(k = e \oplus m|m) = \frac{1}{2^n} \, , \tag{1.63}$$

since k is chosen uniformly at random. Since this holds for any message m,

$$p(e) = \sum_m p(m)p(e|m) = \frac{1}{2^n} \, . \tag{1.64}$$

Applying Bayes' rule we get that

$$p(m|e) = \frac{p(m,e)}{p(e)} = \frac{p(e|m)p(m)}{p(e)} = p(m)\ , \tag{1.65}$$

independent of m. Thus $p(m|e) = p(m)$ and the scheme is perfectly secure.

Note that the argument crucially relies on the key being uniformly distributed and independent from the eavesdropper, a condition that has to be treated with care! In Chapter **??** 4 we will introduce a method called *privacy amplification* that can be used to "improve" the quality of a key about which the eavesdropper may have partial information.

Remark 1.5.1. *We note that while the one-time pad is perfectly "secure" according to Shannon's definition, it does not protect against an adversary changing bits in the messages exchanged between Alice and Bob. Indeed, you can verify that for any key k, and any string x,* $\mathrm{Enc}(xm \oplus x, k) = \mathrm{Enc}(m,k) \oplus x$. *What this means is that flipping bits of the ciphertext is equivalent to flipping bits of the plaintext, and there is no way for Bob to detect if such an operation has taken place. This would be an issue for bank transactions, since an adversary could flip the transaction amount in an arbitrary way (without ever learning any information about the amount itself!). For this reason, one-time pads are generally supplemented by checksums or message authentication codes which allow changes to be detected (and corrected). These are well-known classical techniques, and we will not get into them in more detail here.*

There is another way to look at the classical one time pad that brings it much closer to the quantum version we will consider next. Consider the encryption of a single-bit message $m \in \{0,1\}$. Recall that we can represent this message as a pure quantum state $|m\rangle$, or equivalently as the density matrix $|m\rangle\langle m|$. When we apply the XOR operation the result is that the bit m is flipped whenever the key bit $k = 1$, and unchanged if $k = 0$. That is, when $k = 1$ the state is transformed as $|m\rangle \mapsto X\,|m\rangle$, where recall that X is the Pauli bit-flip matrix. In terms of the density matrix, the encryption implements the transformation $|m\rangle\langle m| \mapsto X|m\rangle\langle m|X$. If Alice and Bob choose a random key bit k, then from the point of view of the eavesdropper (who does not have access to k) the state of the message is represented by the probabilistic mixture

$$\rho = \frac{1}{2}|m\rangle\langle m| + \frac{1}{2}X|m\rangle\langle m|X = \frac{\mathbb{I}}{2}\ . \tag{1.66}$$

Note that ρ does *not* depend on m! Whatever m is, we get that $\rho = \frac{\mathbb{I}}{2}$. Since all information that can be gained from receiving the encrypted message is captured in the density matrix ρ, it follows that absolutely no information about m can be gained from intercepting the encryption.

The quantum one-time pad

Let us consider the task of encrypting a qubit, instead of a classical bit [**quantumOneTimePad**, **QOTP2**]. Instead of one key bit, however, it turns out that we require two key bits $k_1 k_2$ to encrypt a single qubit. Indeed, it can be shown that two key bits are *necessary* . An intuition for why more than one key bit is needed is that a quantum encryption scheme should hide information in all possible bases the qubit could be encoded in. In the classical case applying the bit flip operator X allowed us to encrypt any bit expressed in the standard basis. If we are allowed other bases, we should also encrypt a bit encoded in the Hadamard basis. However, in which case $X|+\rangle\langle+|X = |+\rangle\langle+|$ and $X|-\rangle\langle-|X = |-\rangle\langle-|$. In other words, the qubit is unchanged by the "encryption" procedure, and the scheme is completely insecure. Having made this observation, it may even seem miraculous that quantum encryption is at all possible using only a finite amount of key! But it is possible, and we will see how.

The trick for the quantum one-time pad is to apply a bit flip in both bases, standard and Hadamard: as it turns out, considering these two bases only will be sufficient to hide information that may be encoded in *any* basis. To flip in both bases, we aply the unitary operator $X^{k_1}Z^{k_2}$, where $k_1, k_2 \in \{0,1\}$ are two key bits chosen uniformly at random. With this choice of encryption operation, an arbitrary single-qubit ρ is transformed as

$$\rho \mapsto \frac{1}{4} \sum_{k_1,k_2 \in \{0,1\}} X^{k_1} Z^{k_2} \rho Z^{k_2} X^{k_1} \ . \tag{1.67}$$

To see why this works, remember the Bloch sphere representation of ρ and the fact that the Pauli matrices pairwise anti-commute. In particular, applying either $\mathbb{I}$, X, Z or XZ with equal probability to the Pauli matrix X gives

$$\begin{aligned} \frac{1}{4}\big(X + XXX + ZXZ + XZXZX\big) &= \frac{1}{4}\big(X + X - ZZX - XZZXX\big) \\ &= \frac{1}{4}\big(X + X - X - X\big) &&= 0 \ , \end{aligned}$$

where we used the fact that the Pauli matrices are observables (i.e. they are Hermitian and square to identity), and $\{X, Z\} = XZ + ZX = 0$. Refer to Figure 1.1 for a visualization of what happened in this calculation.

Exercise 1.5.1. *Show that similarly, for any $M \in \{X, Z, Y\}$ we have*

$$\frac{1}{4} \sum_{k_1,k_2} X^{k_1} Z^{k_2} M Z^{k_2} X^{k_1} \ = 0. \tag{1.68}$$

■

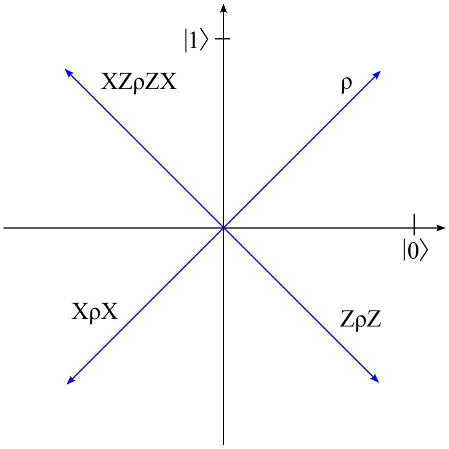

Figure 1.1: A qubit encoded by two key bits: the operations $\mathbb{I}, X, Z, XZ$ are performed on the qubit with equal probability. The resulting mixture of states is the maximally mixed state (represented by the origin of the diagram).

Since any single-qubit state can be written as

$$\rho = \frac{1}{2}\left(\mathbb{I} + v_x X + v_y Y + v_z Z\right) \ , \tag{1.69}$$

we get using linearity and the exercise that for any ρ,

$$\frac{1}{4}\sum_{k_1,k_2} X^{k_1} Z^{k_2} \rho Z^{k_2} X^{k_1} \ = \frac{\mathbb{I}}{2} \ . \tag{1.70}$$

What this equation means is precisely that from the point of view of anyone who does not know k_1, k_2 the bit- and phase-flipped state is completely independent of the input ρ, which means that all information contained in ρ is hidden from the eavesdropper. This leads to the following quantum encryption scheme.

Protocol 2. *The quantum one-time pad is an encryption scheme for qubits. The key $k = (k_1, k_2)$ is chosen uniformly at random in $K = \{0,1\}^2$. To encrypt a qubit in state ρ, Alice applies the unitary operation $X^{k_1}Z^{k_2}$ to ρ. To decrypt, Bob applies the inverse operation $(X^{k_1}Z^{k_2})^\dagger = Z^{k_2}X^{k_1}$.*

The fact that the scheme is correct follows by definition, since the decryption operation is the inverse of the encryption operation. We have argued that security follows from (1.70). While we have not yet given an entirely formal definition of security in the quantum case, we will develop such a definition over the coming chapters, and this will allow us to show that the scheme is indeed “perfectly secure” in the same sense as the classical one-time pad is.

The quanutm one-time pad can easily be extended to n qubits by applying either $\mathbb{I}$, X, Z or XZ on each qubit, depending on two key bits associated with that qubit. This means that to encrypt n qubits, we use $2n$ bits of classical key.

Exercise 1.5.2. *Show that the collection of all (normalized) tensor products of Pauli matrices*

$$P^s = \frac{1}{2^n} X^{s_1} Z^{s_2} \otimes X^{s_3} Z^{s_4} \otimes \ldots \otimes X^{s_{2n-1}} Z^{s_{2n}}$$

with $s \in \{0,1\}^{2n}$ form an orthogonal basis for all linear operators $\mathcal{L}(\mathbb{C}^{2^n}, \mathbb{C}^{2^n})$, i.e. for all $s, t \in \{0,1\}^{2n}$, $\mathrm{tr}[(P^s)^\dagger P^{\hat{s}}] = \delta_{st}$. In particular, any density matrix ρ on n qubits has a unique decomposition of the form

$$\rho = \frac{1}{2^n} \left(\mathbb{I}^{\otimes 2n} + \sum_{s \neq 0} v_s P^s \right) . \tag{1.71}$$

■

Remark 1.5.2. *It would be natural to think that for n-qubit systems as for 1-qubit systems the coefficients v_s associated with density matrices could be characterized by some form of higher-dimensional analogue of the Bloch sphere. This is not true, and much more complicated conditions on the coefficients v_s have to hold for ρ to be a valid quantum state. The Bloch sphere representation is only used for a single qubit, where it forms a useful visualization tool.*

1.6 Important identities for calculations

Trace

Given a matrix M, the trace is given by $\mathrm{tr}(M) = \sum_i M_{ii}$, i.e. the sum of its diagonal elements. The trace operation is cyclic, i.e. for any two matrices M, N, $\mathrm{tr}(MN) = \mathrm{tr}(NM)$.

Density Matrices

If a source prepares a quantum system in the state ρ_x with probability p_x, then the resulting state of the system is given by the density matrix

$$\rho = \sum_x p_x \rho_x. \tag{1.72}$$

Bloch representation of density matrices: any qubit density matrix can be written as

$$\rho = \frac{1}{2}\left(\mathbb{I} + v_x X + v_z Z + v_y Y\right), \tag{1.73}$$

and the Bloch vector $\vec{v} = (x_x, v_y, v_z) \leq 1$ with equality if and only if ρ is pure.

Probability of measurement outcomes on a density matrix

If a quantum state with density matrix ρ is measured in the basis $\{|w_j\rangle\}_j$, then the probabilities of obtaining each outcome $|w_j\rangle$ is given by

$$p_{w_j} = \langle w_j|\,\rho\,|w_j\rangle = \mathrm{tr}(\rho|w_j\rangle\langle w_j|). \tag{1.74}$$

Combining density matrices

For density matrices $\rho_A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $\rho_B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ representing qubits A and B, the joint density matrix is given by

$$\rho_{AB} = \rho_A \otimes \rho_B := \begin{pmatrix} a_{11}\rho_B & a_{12}\rho_B \\ a_{21}\rho_B & a_{22}\rho_B \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}. \tag{1.75}$$

Partial trace

Given a bipartite matrix ρ_{AB}, which can be expressed in a general form:

$$\rho_{AB} = \sum_{ijkl} \gamma_{ij}^{kl} |i\rangle\langle j| \otimes |k\rangle\langle l|, \tag{1.76}$$

the partial trace operation over system A yields the reduced state ρ_B

$$\rho_B = \mathrm{tr}_A(\rho_{AB}) = \sum_{ijk\ell} \gamma_{ij}^{k\ell}\,\mathrm{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle\ell| = \sum_{k\ell}\left(\sum_j \gamma_{jj}^{k\ell}\right)|k\rangle\langle\ell|\,. \quad (1.77)$$

Properties of Pauli Matrices X, Z, Y

For any $S_1, S_2 \in \{X, Y, Z\}$, $\{S_1, S_2\} = 2\delta_{S_1 S_2}\mathbb{I}$ where the anti-commutator is $\{A, B\} = AB + BA$. This implies the following

1. Zero trace: $\mathrm{tr}(S_1) = 0$.
2. Orthogonality: $\mathrm{tr}(S_1^\dagger S_2) = 0$.
3. Unitary: $S_1^\dagger S_1 = S_1 S_1^\dagger = \mathbb{I}$.
4. Squared to identity: $S_1^2 = \mathbb{I}$.