Lecture Notes

Quantum Cryptography Week 1: Quantum tools and a first protocol

Contents

In this course you will learn about the basics of quantum communication and quantum cryptography. Unlike large scale quantum computers, both technologies are already implemented today. Yet it remains a grand challenge to do quantum communication and cryptography over long distances. This week we will learn about a very simple quantum protocol: we will encrypt quantum states! Yet, to prepare our entry into quantum communication and cryptography, we first need to learn a little more about quantum information. Even if you did not follow week 0, we recommend downloading the lecture notes for week 0 for notations and conventions used here.

1.1 Probability notation

Before we start we recall standard notation of classical probability theory which we use throughout these lecture notes. There are many good textbooks and online resources on probability theory available, such as [Kel94; Ros10], and we refer you to any of them for additional background.

Consider a discrete random variable X taking values in some alphabet $\mathscr{X}$ of size n. We write $P_X(\cdot)$ for the distribution of X, and $|X|$ for the size of the alphabet of X. The notation $P_X(x)$ denotes the probability that the random variable takes on a specific symbol $x \in \mathscr{X}$. When the distribution is clear from context, we use the shorthands $p_x = p(x) = P(X = x) = P_X(x)$. It will be useful to remember that a probability distribution $P_X(\cdot)$ is specified by non-negative probability values, i.e. $\forall x \in \mathscr{X}$, $P_X(x) \geq 0$. Furthermore, X should be normalized, which means $\sum_{x\in\mathscr{X}} P_X(x) = 1$.

■ **Example 1.1.1** Let $\mathscr{X} = \{1,2,3,4,5,6\}$ correspond to the faces of a 6 sided die. If the die is fair, i.e., all sides have equal probability of occuring then $P_X(x) = 1/6$ for all $x \in \mathscr{X}$. Using the shorthands, this reads $p_x = p(x) = 1/6$. The size of the alphabet is given by $|X| = 6$. ■

A random variable X can be correlated with another random variable Y. This means that they have a joint distribution, $P_{XY}(x,y)$, that is not necessarily a product, that is, $P_{XY}(x,y) \neq P_X(x)P_Y(y)$ in general. 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, conditional on the event that Y takes on the value y. As before, we will generally use the following shorthands when it is clear which random variable we refer to

$$p_{x|y} = p(x|y) = P(X = x|Y = y) = P_{X|Y}(x|y). \tag{1.1}$$

As you know from your probability class, Bayes rule relates this conditional probability to the joint probabilities. Since $P_{XY}(x,y) = P_X(x)P_{Y|X}(y|x) = P_Y(y)P_{X|Y}(x|y)$ we have

$$P_{X|Y}(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)}\,, \tag{1.2}$$

whenever $P_Y(y) > 0$ [1].

■ **Example 1.1.2** Let's consider the fair die above, and an unfair die which always rolls a "6". That is, $\mathscr{X} = \{1,2,3,4,5,6\}$ in which $P_X(6) = 1$ and $P_X(x) = 0$ for $x \neq 6$. Let Y now refer to the choice of a fair, or unfair die. Suppose that we choose to roll the unfair or fair die with equal probability. That is $\mathscr{Y} = \{$"fair", "unfair"$\}$ where $P_Y(\text{fair}) = 1/2$ and $P_Y(\text{unfair}) = 1/2$. We thus have $P_{X|Y}(x|\text{fair}) = 1/6$ and $P_{X|Y}(6|\text{unfair}) = 1$ and $P_{X|Y}(x|\text{unfair}) = 0$ for $x \neq 6$. ■

Exercise 1.1.1 Compute the joint probability $P_{XY}(x,y)$ for the example of choosing a fair or unfair die. ■

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

Exercise 1.1.2 Suppose now that we choose to roll the fair or unfair die with probability $P_Y(\text{fair}) = P_Y(\text{unfair}) = 1/2$, but don't tell you which one it is. However, I show you the outcome X of the die roll. That is, I have Y and you have X. Suppose that $X = 3$. What is the most likely die? I.e., is it more likely that $Y = \text{fair}$ or $Y = \text{unfair}$? How about $X = 6$? ■

1.2 Density matrices

Let us start by investigating a more general formalism for writing down quantum states. There are two motivations for doing so. Let's start with the basic question of how to write down the state of one of several qubits. To this end, imagine we have two quantum systems A and B. For example, A and B are two qubits in a joint state $|\psi\rangle_{AB}$ and we want to know the state of qubit A. If the joint state is $|\psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B$, that is, it is obtained by taking the tensor product of qubit A in the state $|\psi\rangle_A$ and qubit B in the state $|\psi\rangle_B$, then the answer seems clear: A is simply in the state $|\psi\rangle_A$. However, if you took week 0, you may remember that some bipartite quantum states $|\psi\rangle_{AB}$, defined over two systems A and B, can be defined as superpositions of tensor products, 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 $|\text{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}|0\rangle_A|0\rangle_B + \frac{1}{\sqrt{2}}|1\rangle_A|1\rangle_B$. For such states we *cannot* express $|\psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B$, that is as a tensor product of two states $|\psi\rangle_A$ on A and $|\psi\rangle_B$ on B. It is thus unclear how we could express the state of A without making any reference to B. Such a description should still be possible: after all, the state does exist! If it doesn't fit in our formalism of states as vectors it must mean the formalism is incomplete, and we need to find a mathematical generalization for it.

The second motivation for a more general description arises from a situation in which a probabilistic process, for example a measurement, prepares different states with some probability. Suppose we encounter a situation in which we had either a state $|\psi_1\rangle$ with some probability p_1, or a state $|\psi_2\rangle$ with probability p_2. To express the state accurately, we have to take into account both states and probabilities $\{|\psi_i\rangle, p_i\}_i$. Can we somehow write down the proper mathematical description of the state created by such a process?

1.2.1 Introduction

The answer to these questions lies in the so-called density matrix formalism. To start with, let us write down the quantum state $|\psi\rangle$ of a single system as a matrix $\rho = |\psi\rangle\langle\psi|$. Note that this is a rank-1 matrix, it has precisely 1 non-zero eigenvalue (equal to 1) with associated eigenstate $|\psi\rangle$.

■ **Example 1.2.1** Consider the following matrices corresponding to $|0\rangle$ and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$

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

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

■

How does writing down states as matrices help us to resolve the questions above? To see how, let us first consider the second motivation for a more general description. In particular, let us consider the case where someone prepares 2 possible states $|\psi_1\rangle$ and $|\psi_2\rangle$ with equal probability $p_1 = p_2 = 1/2$. Clearly, a superposition is not the correct description: The state really is in precisely one of the two states, with probability $1/2$ each. Indeed, the preparer knows the identity of the state. If the identity of the state is not known, however, how can we write down the resulting state? It turns out that we can describe the state of the resulting system as a *mixture* between $|\psi_1\rangle$ and $|\psi_2\rangle$.

For equal probabilities, this mixture becomes

$$\rho = \frac{1}{2}|\psi_1\rangle\langle\psi_1| + \frac{1}{2}|\psi_2\rangle\langle\psi_2| \,. \tag{1.5}$$

We also call such a ρ a *density matrix*. In general, if a source prepares the state $|\psi_x\rangle$ with probability p_x, the resulting system will be in the state

$$\rho = \sum_x p_x |\psi_x\rangle\langle\psi_x| \,. \tag{1.6}$$

Why would this be a good description? Let's consider what happens if we measure in the standard basis. If the system would actually be in the state $|\psi_j\rangle$, then we would expect the probabilities of outcomes to be

$$q_{0|j} = |\langle 0||\psi_j\rangle|^2 = \langle 0||\psi_j\rangle\langle\psi_j||0\rangle \,, \tag{1.7}$$
$$q_{1|j} = |\langle 1||\psi_j\rangle|^2 = \langle 1||\psi_j\rangle\langle\psi_j||1\rangle \,. \tag{1.8}$$

If state $|\psi_j\rangle$ is prepared with probability p_j, then we would expect the outcome probabilities to be

$$q_0 = \sum_j p_j q_{0|j} \,, \tag{1.9}$$
$$q_1 = \sum_j p_j q_{1|j} \,. \tag{1.10}$$

Let us expand one of these terms to relate to the density matrix formalism. We have

$$q_0 = \sum_j p_j q_{0|j} = \sum_j p_j \langle 0||\psi_j\rangle\langle\psi_j||0\rangle = \langle 0| \left(\sum_j p_j |\psi_j\rangle\langle\psi_j| \right) |0\rangle = \langle 0|\rho|0\rangle \,. \tag{1.11}$$

The density matrix ρ thus accurately reflects what we would intuitively expect from the probabilities of measurement outcomes.

■ **Example 1.2.2** If a source prepares quantum states in a probabilistic manner, i.e. it prepares the quantum state ρ_x with probability p_x, then the resulting *density matrix* is given by

$$\rho = \sum_x p_x \rho_x \,. \tag{1.12}$$

The set of probabilities and density matrices $\mathscr{E} = \{(p_x, \rho_x)\}_x$ is also called an *ensemble* of states. Note that the case where the source prepares pure states is a special case with $\rho_x = |\psi_x\rangle\langle\psi_x|$ and $p_x = 1$ for a single x. ■

■ **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

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \,. \tag{1.13}$$

■

■ **Example 1.2.4** Superposition is not the same as a mixture. Intuitively, the difference is that a mixture is an inherently classical mixture: there is a process that prepares one *or* the other with some probability. In contrast, a state in a superposition is one *and* the other. To see the difference,

let us consider mixing or creating a superposition of $|0\rangle$ and $|1\rangle$. Consider a source that prepares the state $|0\rangle$ and $|1\rangle$ with probabilities $p_0 = p_1 = 1/2$. Suppose we measure the resulting state

$$\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \mathbb{I}/2 \,, \tag{1.14}$$

where

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{1.15}$$

in the Hadamard basis $\{|+\rangle, |-\rangle\}$ with

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \,, \tag{1.16}$$

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \,. \tag{1.17}$$

We have that the probabilities of outcomes are given by

$$q_+ = \langle +|\rho|+\rangle = \frac{1}{2} \,, \tag{1.18}$$

$$q_- = \langle -|\rho|-\rangle = \frac{1}{2} \,. \tag{1.19}$$

In contrast, consider now the superposition $|+\rangle$. Measuring $|+\rangle$ in the Hadamard basis, results in $q_+ = 1$ and $q_- = 0$. This illustrates a fundamental difference between mixtures and superpositions. ■

Exercise 1.2.1 If $|\Psi\rangle$ is an n-qubit quantum state, what are the dimensions of the density matrix $|\Psi\rangle\langle\Psi|$? ■

R It is crucial to note that unlike in the case of classical probability distributions, the same density matrix can be obtained from different ensembles. A simple example is provided by the operator

$$\rho = \frac{\mathbb{I}}{2} \,, \tag{1.20}$$

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

$$\frac{\mathbb{I}}{2} = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{1}{2}(|+\rangle\langle +| + |-\rangle\langle -|) \,. \tag{1.21}$$

1.2.2 Some mathematical definitions

To formally define density matrices and their properties, we recall some important notions from linear algebra. The first term we introduce is the *linear operator*, which is, in our context, just a fancy name to express matrices. Such a term highlights the idea that a matrix maps one vector to another. It can hence be thought of as an operation performed on vectors, which - since matrix multiplication is linear - will be a linear operation. We will hence use matrix and operator interchangeably. To make sense of the quantum literature, however, the following definitions will be useful.

Definition 1.2.1 — Linear operator. Consider a d-dimensional complex vector space $\mathbb{C}^d$. A *linear* operator $L : \mathbb{C}^d \to \mathbb{C}^{d'}$ can be represented as a $d' \times d$ matrix,

$$L = \begin{pmatrix} L_{11} & L_{12} & \cdots & L_{1d} \\ L_{21} & \ddots & \ddots & L_{2d} \\ \vdots & \ddots & \ddots & \vdots \\ L_{d'1} & L_{d'2} & \cdots & L_{d'd} \end{pmatrix}, \tag{1.22}$$

where each element $L_{ij} \in \mathbb{C}$. The set of linear operators is denoted $\mathscr{L}(\mathbb{C}^d, \mathbb{C}^{d'})$.

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

The spectral theorem states that any Hermitian operator $M \in \mathscr{L}(\mathbb{C}^d, \mathbb{C}^d)$ 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, i.e. $\lambda_i \geq 0$. This condition is often denoted as $M \geq 0$.

Exercise 1.2.2 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. ■

An important operation on matrices is the *trace*. We already saw in Week 0 that we can express it simply as the sum of the diagonal elements. As such, the trace is a linear map which takes any matrix to a complex number. It will sometimes be 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 $M \in \mathscr{L}(\mathbb{C}^d, \mathbb{C}^d)$ is defined as

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

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! An important property of the trace is that it is *cyclic*:

Exercise 1.2.3 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

$$\langle i|M|i\rangle = \mathrm{tr}(\langle i|M|i\rangle) = \mathrm{tr}(M|i\rangle\langle i|). \tag{1.23}$$

It is however worth noting that in general, a non-cyclic permutation of the matrices do not preserve the trace. More precisely, for matrices M, N, P, in general

$$\mathrm{tr}(MNP) \neq \mathrm{tr}(NMP). \tag{1.24}$$

■

1.2.3 Density matrices and their properties

Given the discussion above we are motivated to take the density matrix ρ as a more general description of quantum states. Before was can make this formally precise, let us first investigate when some matrix ρ would actually be considered a valid density matrix, that is, a description of a quantum state. It turns out that there are two necessary (and sufficient) properties in order for a density matrix to represent a valid quantum state: it 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 into its eigenvalues $\{\lambda_j\}_j$ and corresponding eigenvectors $\{|v_j\rangle\}_j$ as

$$\rho = \sum_j \lambda_j |v_j\rangle\langle v_j| \tag{1.25}$$

where the vectors $|v_j\rangle$ are orthonormal. Let us imagine that we measure ρ in some other orthonormal basis $\{|w_k\rangle\}_k$. Thinking about a process that prepares a certain state $|v_j\rangle\langle v_j|$ with probability λ_j, we could imagine that we measure just $|v_j\rangle$ in that basis. We know that in this case, the probability of obtaining measurement outcome k (conditioned on the preparation being in state $|v_j\rangle\langle v_j|$) is given by

$$q_{k|j} = |\langle v_j|w_k\rangle|^2 = \langle w_k||v_j\rangle\langle v_j||w_k\rangle \ . \tag{1.26}$$

Hence the probability of obtaining outcome k when measuring ρ should be given by

$$q_k = \sum_j \lambda_j q_{k|j} = \langle w_k| \left(\sum_j \lambda_j |v_j\rangle\langle v_j| \right) |w_k\rangle = \langle w_k|\rho|w_k\rangle \ . \tag{1.27}$$

Note that we must have $q_k \geq 0$ and $\sum_k q_k = 1$. By imagining that we measure ρ in its eigenbasis, that is, $|w_j\rangle = |v_j\rangle$, it is easy to see that $\lambda_j \geq 0$, that is, ρ is a *positive semidefinite* matrix. We also have $\mathrm{tr}(\rho) = 1$, since

$$1 = \sum_j q_j = \sum_j \lambda_j \mathrm{tr}(|v_j\rangle\langle v_j|) = \mathrm{tr}(\rho) \ . \tag{1.28}$$

This motivates the following definition of a density matrix, which is the most general way to describe the state of a quantum system.

Definition 1.2.5 — Density matrix. Consider a quantum system with state space $\mathbb{C}^d$. A *density matrix*, commonly denoted as ρ, is a linear operator $\rho \in \mathscr{L}(\mathbb{C}^d, \mathbb{C}^d)$ such that:
1. $\rho \geq 0$, and
2. $\mathrm{tr}(\rho) = 1$.
If $\mathrm{rank}(\rho) = 1$, then ρ is called a *pure* state, otherwise ρ is *mixed*.

Let us also summarize the rule for computing outcome probabilities for measuring a quantum system described by the density matrix ρ motivated by our discussions.

Definition 1.2.6 — Measuring a density matrix in a basis. Consider a quantum system in the state ρ. Measuring ρ in the basis $\{|b_j\rangle\}_j$ results in outcome j with probability

$$q_j = \langle b_j|\rho|b_j\rangle \ . \tag{1.29}$$

1.2.4 Bloch representation for one qubit mixed states

In week 0, we saw that one qubit states have a nice graphical representation in terms of vectors on the Bloch sphere. In particular, any pure quantum state can be described by a *Bloch vector*

$\vec{r} = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)$. Is this the same for mixed states? The answer to this turns out to be yes! Concretely, we can write any one qubit density matrix as

$$\rho = \frac{1}{2}(\mathbb{I} + v_x X + v_z Z + v_y Y) \ , \tag{1.30}$$

where X, Y, Z are the Pauli matrices you have encountered in Week 0, and if ρ is pure, then the vector $\vec{v} = (v_x, v_y, v_z)$ is precisely the Bloch vector $\vec{r}$ that you already know! 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 $\|\vec{v}\|_2 \leq 1$.

Note that the matrices $\mathscr{S} = \{\mathbb{I}, X, Z, Y\}$ 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,

$$\mathrm{tr}(X^\dagger Y) = \mathrm{tr}(X^\dagger Z) = \mathrm{tr}(X^\dagger \mathbb{I}) = 0 \ , \tag{1.31}$$

and similarly for all other pairs of matrices.

Exercise 1.2.4 Using the orthogonality condition (1.31), show that

$$|0\rangle\langle 0| = \frac{1}{2}(\mathbb{I} + Z) \ , \tag{1.32}$$

$$|1\rangle\langle 1| = \frac{1}{2}(\mathbb{I} - Z) \ , \tag{1.33}$$

■

Exercise 1.2.5 Use the fact that all matrices $M, N \in \mathscr{S}$ with $M \neq N$ anti-commute, i.e., $\{M, N\} = MN + NM = 0$ to show that $\mathrm{tr}(MN) = 0$ whenever $M \neq N \in \mathscr{S}$. ■

1.3 Combining density matrices

If we have two quantum systems A and B, described by density matrices ρ_A and ρ_B, how can we write down the joint state ρ_{AB}? We saw in Week 0 that two pure quantum states which can be represented by vectors $|v_1\rangle \in \mathbb{C}^{d_1}, |v_2\rangle \in \mathbb{C}^{d_2}$ 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 concept of the tensor product. Let us start with the simple case where ρ_A, ρ_B are 2×2-dimensional matrices,

$$\rho_A \otimes \rho_B = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \otimes \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} = \begin{pmatrix} m_{11}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} & m_{12}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \\ m_{21}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} & m_{22}\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \end{pmatrix} \tag{1.34}$$

$$= \begin{pmatrix} m_{11}n_{11} & m_{11}n_{12} & m_{12}n_{11} & m_{12}n_{12} \\ m_{11}n_{21} & m_{11}n_{22} & m_{12}n_{21} & m_{12}n_{22} \\ m_{21}n_{11} & m_{21}n_{12} & m_{22}n_{11} & m_{22}n_{12} \\ m_{21}n_{21} & m_{21}n_{22} & m_{22}n_{21} & m_{22}n_{22} \end{pmatrix} . \tag{1.35}$$

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

$$\rho_{AB} = \rho_A \otimes \rho_B = \begin{pmatrix} 1\cdot\rho_B & 0\cdot\rho_B \\ 0\cdot\rho_B & 0\cdot\rho_B \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{1.36}$$

This definition easily extends to larger matrices as follows:

Definition 1.3.1 — Tensor product. Consider any $d' \times d$ matrix ρ_A and $k' \times k$ matrix ρ_B,

$$\rho_A = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1d} \\ m_{21} & \ddots & \ddots & m_{2d} \\ \vdots & \ddots & \ddots & \vdots \\ m_{d'1} & m_{d'2} & \cdots & m_{d'd} \end{pmatrix}, \qquad \rho_B = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1k} \\ n_{21} & \ddots & \ddots & n_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ n_{k'1} & n_{k'2} & \cdots & n_{k'k} \end{pmatrix}. \tag{1.37}$$

The tensor product of these matrices is given by a $d'k' \times dk$ matrix,

$$\rho_{AB} = \rho_A \otimes \rho_B = \begin{pmatrix} m_{11}B & m_{12}B & \cdots & m_{1d}B \\ m_{21}B & \ddots & \ddots & m_{2d}B \\ \vdots & \ddots & \ddots & \vdots \\ m_{d'1}B & m_{d'2}B & \cdots & m_{d'd}B \end{pmatrix}. \tag{1.38}$$

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

■ **Example 1.3.1** 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

$$\rho_A \otimes \rho_B = \frac{1}{8}\begin{pmatrix} 1 & -i & 1 & -i & 0 & 0 \\ i & 1 & i & 1 & 0 & 0 \\ 1 & -i & 2 & -2i & 1 & -i \\ i & 1 & 2i & 2 & i & 1 \\ 0 & 0 & 1 & -i & 1 & -i \\ 0 & 0 & i & 1 & i & 1 \end{pmatrix}, \tag{1.39}$$

and

$$\rho_B \otimes \rho_A = \frac{1}{8}\begin{pmatrix} 1 & 1 & 0 & -i & -i & 0 \\ 1 & 2 & 1 & -i & -2i & -i \\ 0 & 1 & 1 & 0 & -i & -i \\ i & i & 0 & 1 & 1 & 0 \\ i & 2i & i & 1 & 2 & 1 \\ 0 & i & i & 0 & 1 & 1 \end{pmatrix} \neq \rho_A \otimes \rho_B. \tag{1.40}$$

■

1.4 Classical-quantum states

Throughout quantum cryptography, we often find ourselves in a situation in which the honest parties have some classical information X about which an adversary - like an eavesdropper Eve - may hold some quantum information Q. It is worth thinking about that the joint states ρ_{XQ} have a very special structure.

1.4.1 Classical states

As a first step, let us first pause to think about what it means that X is "classical information". To this end, it is interesting to note that it is possible to write a probability distribution over classical strings x in terms of density matrices. Suppose that we have a classical probability distribution over symbols from the alphabet $\mathscr{X} = \{0,\dots,d-1\}$, where p_x denotes the probability of observing symbol x. Identifying classical bits (or indeed numbers) with elements of the standard basis $\{|0\rangle,\dots,|d-1\rangle\}$, we can express a source preparing each of the possible states $|x\rangle$ with probability p_x by the mixture

$$\rho = \sum_{x=0}^{d-1} p_x |x\rangle\langle x| \,. \tag{1.41}$$

Note that ρ is a density matrix which has the probabilities p_x on the diagonal and is otherwise zero. As such, ρ is just another way to express the probability distribution p_x. Indeed, you may want to check that measuring ρ in the standard basis results precisely in obtaining outcome "x" with probability p_x.

Definition 1.4.1 — Classical state. Consider a system X with state space $\mathbb{C}^d$, and let $\{|x\rangle\}_{x=0}^{d-1}$ denote the standard basis for $\mathbb{C}^d$. A system X is in a classical state, or *c-state*, when the corresponding density matrix ρ_X is diagonal in the standard basis of the state space of X, i.e. ρ_X has the form

$$\rho = \sum_{x=0}^{d-1} p_x |x\rangle\langle x| \tag{1.42}$$

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

In quantum cryptography, we will often encounter states which are partially classical, and partially quantum. Suppose we prepare the following states for Alice and Bob. With probability $1/2$ we prepare $|0\rangle\langle 0|_X \otimes \rho_0^Q$ with $\rho_0^Q = \frac{\mathbb{I}_Q}{2}$, and with probability $1/2$ we prepare $|1\rangle\langle 1|_X \otimes \rho_1^Q$ with $\rho_1^Q = |+\rangle\langle +|$. The joint state is a *classical quantum state*, or cq-state of the form

$$\rho_{XQ} = \frac{1}{2} \sum_{x\in\{0,1\}} |x\rangle\langle x|_X \otimes \rho_x^Q \,. \tag{1.43}$$

Note that in this case Alice knows which state Bob is given. However, as we will see later, Bob cannot learn which x Alice holds with certainty. (Intuitively, the reason is that, while Alice's states are orthonormal, Bob's states have "overlap" and are not perfectly distinguishable.)

Definition 1.4.2 A classical-quantum state, or simply called a *cq-state* takes the form

$$\rho_{XQ} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_x^Q \,. \tag{1.44}$$

That is, it consists of a classical register X and a quantum register Q. If Q is absent, then ρ_X is simply a classical state.

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 attacker 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 use letters X, Y, Z to denote such classical registers, and reserve the other letters for quantum information.

1.5 General measurements

So far we have only been measuring quantum states 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. On the one hand, in quantum information theory certain tasks, such as the task of discriminating between multiple states, can be solved more efficiently using these generalized measurements. On the other hand, taking an adversarial viewpoint, in quantum cryptography it is essential that we prove security for the most general kind of attack, including all measurements that an attacker could possibly make!

1.5.1 POVMs

If we are only interested in the probabilities of measurement outcomes - but not what happens after the measurement - 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. It can be defined as follows.

Definition 1.5.1 — POVM. A POVM on $\mathbb{C}^d$ is a set of positive semidefinite operators $\{M_x\}_{x\in\mathscr{X}}$ such that

$$\sum_x M_x = \mathbb{I}_{\mathbb{C}^d} \ . \tag{1.45}$$

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

$$p_x = \mathrm{tr}(M_x\rho) \ . \tag{1.46}$$

■ **Example 1.5.1** 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 outcome x is simply given by $|\alpha_x|^2$. Let's see how this can be formulated 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 $\sum_x M_x = \mathbb{I}$ ($\{|x\rangle\}$ is a basis), as required. We can then 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 1.5.2** Consider a distribution (p_x) and the 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 previous example, we obtain outcome x with probability

$$\mathrm{tr}(|x\rangle\langle x|\rho) = \langle x|\rho|x\rangle = p_x. \tag{1.47}$$

Thus ρ indeed captures the classical distribution given by the probabilities p_x. ■

You may wonder what happens to a quantum state after a generalized measurement has been performed. For the case of measuring in a basis, the answer is simple: the state 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. Intuitively the reason is because such a measurement may not fully collapse the state (the post-measurement state may not be pure), and as a consequence there remains the flexibility to apply an arbitrary unitary on the post-measurement state, without affecting the outcome probabilities.

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

Definition 1.5.2 — 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 $A_x \in \mathscr{L}(\mathbb{C}^d, \mathbb{C}^{d'})$ 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.

1.5.2 Generalized measurements

The most general form to write down a quantum measurement is thus 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 also compute the post-measurement states.

Definition 1.5.3 — 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 preformed 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 the standard basis, with POVM elements $M_x = |x\rangle\langle x|$ and Kraus decomposition $A_x = M_x = |x\rangle\langle x|$, the post-measurement state as defined above is precisely the basis state associated to the measurement outcome. Note that since a POVM does not 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, 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.

Definition 1.5.4 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 $A_x = \Pi_x$. The probability q_x of observing measurement outcome x can then 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 1.5.3** 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

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

Now suppose we 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

$$q'_{\text{even}} = \text{tr}(\Pi_{\text{even}}\rho) = \langle 00|\rho|00\rangle + \langle 11|\rho|11\rangle \ , \tag{1.48}$$

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

$$\rho'_{|\text{even}} = \Pi_{\text{even}}\rho\Pi_{\text{even}} \ . \tag{1.49}$$

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 measurement should report the outcome "even" with probability 1, and you can check 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 main 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, as happens when measuring in a basis. ■

Exercise 1.5.1 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? ■

1.6 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 $\mathbb{C}_A^{d_A} \otimes \mathbb{C}_B^{d_B}$, 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?

1.6.1 An operational viewpoint

The operation that takes us from ρ_{AB} to ρ_A is called the *partial trace*. It can be given a purely mathematical description that we will give below. However, before that, let's try to think about the problem from an operational point of view. First, 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. A slightly more complicated case would be when $\rho_{AB} = \sum_i p_i \rho_i^A \otimes \rho_i^B$ is a mixture of tensor products (we will later see this is called a "separable state"); in this case it would certainly be natural to say that Alice's state is ρ_i^A with probability p_i, i.e. $\rho_A = \sum_i p_i \rho_i^A$.

How do we deal with a general ρ? The idea is to *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 $A_x = M_x$ as well. By definition the post-measurement states are given by

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

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.

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 and even 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 is unchanged. We can thus describe it as "with probability $q_x = \mathrm{Tr}(M_x \rho_{AB})$, Alice's state is the A part of $\rho_{|x}^{AB}$, i.e.

$$\rho_A = \sum_x q_x \frac{\left((\mathbb{I} \otimes \langle u_x|)\rho_{AB}(\mathbb{I} \otimes |u_x\rangle)\right)_A}{\mathrm{Tr}\left((\mathbb{I} \otimes |x\rangle\langle x|)\rho_{AB}\right)} = \sum_x (\mathbb{I} \otimes \langle u_x|)\rho_{AB}(\mathbb{I} \otimes |u_x\rangle). \tag{1.50}$$

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 whole 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 the subtlety.)

Exercise 1.6.1 Verify that the state ρ_A defined in Eq.(1.50) 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 1.6.1** Consider the example of the EPR pair

$$|\mathrm{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{1.51}$$

Writing this as a density operator we have

$$\rho_{AB} = |\mathrm{EPR}\rangle\langle\mathrm{EPR}|_{AB} = \frac{1}{2}\left(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\right). \tag{1.52}$$

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

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

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! ∎

1.6.2 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 1.6.1 — Partial Trace. Consider a general state

$$\rho_{AB} = \sum_{ijk\ell} \gamma_{ij}^{k\ell} |i\rangle\langle j|_A \otimes |k\rangle\langle \ell|_B, \tag{1.53}$$

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

$$\rho_A = \mathrm{tr}_B(\rho_{AB}) = \sum_{ijk\ell} \gamma_{ij}^{k\ell} |i\rangle\langle j| \otimes \mathrm{tr}(|k\rangle\langle \ell|) = \sum_{ij}\left(\sum_k \gamma_{ij}^{kk}\right)|i\rangle\langle j| \,. \tag{1.54}$$

Similarly, the partial trace over A becomes

$$\rho_B = \mathrm{tr}_A(\rho_{AB}) = \sum_{ijk\ell} \gamma_{ij}^{k\ell} \mathrm{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle \ell| = \sum_{k\ell}\left(\sum_j \gamma_{jj}^{k\ell}\right)|k\rangle\langle \ell| \,. \tag{1.55}$$

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

∎ **Example 1.6.2** 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

$$\begin{aligned} \mathrm{tr}_B(\rho_{AB}) = \frac{1}{2}\big(&|0\rangle\langle 0| \otimes \mathrm{tr}(|0\rangle\langle 0|) + |0\rangle\langle 1| \otimes \mathrm{tr}(|0\rangle\langle 1|) \\ &+ |1\rangle\langle 0| \otimes \mathrm{tr}(|1\rangle\langle 0|) + |1\rangle\langle 1| \otimes \mathrm{tr}(|1\rangle\langle 1|)\big) \,. \end{aligned} \tag{1.56}$$

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

$$\mathrm{tr}_B(\rho_{AB}) = \frac{1}{2}\left(|0\rangle\langle 0| + |1\rangle\langle 1|\right) = \frac{\mathbb{I}}{2} \,. \tag{1.57}$$

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

$$\mathrm{tr}_A(\rho_{AB}) = \frac{\mathbb{I}}{2}\,. \tag{1.58}$$

■

Exercise 1.6.2 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 1.6.3** We can now see that performing a unitary operation on A has no effect on the state of B, i.e., it does not change ρ_B.

$$(U_A \otimes \mathbb{I}_B)\rho_{AB}(U_A \otimes \mathbb{I}_B)^\dagger = \sum_{ijk\ell} \gamma_{ij}^{k\ell} U_A|i\rangle\langle j|_A U_A^\dagger \otimes |k\rangle\langle\ell|_B. \tag{1.59}$$

Computing again the partial trace we have

$$\begin{aligned}
\mathrm{tr}_A(U_A \otimes \mathbb{I}_B \rho_{AB} U_A^\dagger \otimes \mathbb{I}_B) &= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(U_A|i\rangle\langle j|U_A^\dagger) \otimes |k\rangle\langle\ell| && (1.60)\\
&= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(|i\rangle\langle j|U_A^\dagger U_A) \otimes |k\rangle\langle\ell| && (1.61)\\
&= \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle\ell| && (1.62)\\
&= \sum_{k\ell} \left(\sum_j \gamma_{jj}^{k\ell} \right) |k\rangle\langle\ell| = \rho_B\,. && (1.63)
\end{aligned}$$

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

1.7 Secure message transmission

The first cryptographic challenge that we will consider is the one of secure message transmission. Here, our protagonists Alice and Bob want to protect their communication from the prying eyes of an eavesdropper Eve. Alice and Bob are always honest, and Eve is the *adversary* (sometimes also called *eavesdropper*. Alice and Bob have control over their secure labs that Eve cannot peek into. However, Eve has access to the communication channel connecting Alice and Bob.

The most fundamental (and also the most secure) method that Alice and Bob can use to transmit their messages securely requires them to use a *key* to encode the message. It is assumed that the key is known to both Alice and Bob, but is private to them: Eve has no information about the key. For this reason we call cryptosystems such as the one we're about to discover *private-key* cryptosystems. Today we investigate how such secret key can be used. In later weeks we will use quantum information to come up with the key!

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

Let us assume that Alice and Bob share a classical key k that is unknown to the eavesdropper, in the sense that we will make precise later. For the moment, let us take the intuitive definition that Eve doesn't know the key if she is completely uncorrelated from the key, and $p(k) = 1/|K|$ for $|K|$ possible keys, i.e. every key is equally likely. A mathematical framework for the description of transferring secret messages was first developed in [Sha49]. Any encryption scheme consists of some 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 also called the plaintext, and e the ciphertext. We will also need a decryption function $\mathrm{Dec}(k,e) = m$ that takes the key k and the cipertext e back to the plaintext.

Definition 1.7.1 An encryption scheme $(\mathrm{Enc}, \mathrm{Dec})$ is *secret*, or *secure* if and only if for all prior distributions $p(m)$ over messages, and all messages m, we have

$$p(m) = p(m|e), \tag{1.64}$$

where $e = \mathrm{Enc}(k, m)$.

In other words we call an encryption scheme secret/secure whenever an eavesdropper Eve who may have intercepted the ciphertext e gains no additional knowledge about the message m than she would have without the ciphertext e. That is, 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!

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. At this point, you are probably objecting since surely this would not be very useful! How could Bob hope to learn m, if e has nothing to do with m? The second condition that an encryption scheme has to satisfy is thus that it is *correct*.

Definition 1.7.2 An encryption scheme $(\mathrm{Enc}, \mathrm{Dec})$ is *correct* if and only if for all possible messages m, and all possible keys k, we have $m = \mathrm{Dec}(k, \mathrm{Enc}(k, m))$.

Again it would be easy to find an encryption scheme that is "just" correct: Alice simply sends $e = m$ to Bob. Again, you are possibly objecting, since Eve can now read all messages and this is precisely what we wanted to prevent!

The art of cryptography is to design protocols that are *both* correct *and* secure simultaneously. In almost all situations, it will be easy to be correct, and easy to be secure, but the real challenge arises when we want to combine both conditions.

The secret key we assumed Alice and Bob share will be the essential ingredient required to achieve an encryption scheme that is both correct and secret. Is a key really needed? As it turns out, not only it is needed but in fact we will need just as many keys as there are possible messages. A message is called possible if $p(m) > 0$. Let us establish this fact in a lemma, due to Shannon:

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 scheme using less keys, i.e., $|K| < |M|$. We will show that such a scheme cannot be secure. Consider an eavesdropper who has intercepted the ciphertext e. She could then compute

$$\mathscr{S} = \{\hat{m} \mid \exists k, \hat{m} = \mathrm{Dec}(k, e)\} \ , \tag{1.65}$$

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 $|\mathscr{S}|$ of this set is $|\mathscr{S}| \leq |K|$, since for each possible key k we get at most one message $\hat{m}$. Since $|K| < |M|$, we thus have $|\mathscr{S}| < |M|$. This means that there exists at least one message m such that $m \notin \mathscr{S}$, and hence $p(m|e) = 0$. There is no key which could give this message, so the eavesdropper learns that the message cannot have been m, but one of the other messages instead! Since a message is possible precisely when $p(m) > 0$, we thus have $0 = p(m|e) \neq p(m) > 0$, which violates the security condition. We conclude that the scheme can only be secure if $|K| \geq |M|$. ■

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 shall explore that in the next section.

1.8 The (quantum) one-time pad

Let us consider possibly the simplest scheme to encrypt messages. It is known as the one-time pad, and offers excellent security — we will learn precisely why soon!

1.8.1 The classical one-time pad

Imagine that Alice (the sender) wants to send a secret message m to Bob (the receiver), where we will take $m \in \{0,1\}^n$ to be an n-bit string. Let us furthermore imagine that Alice and Bob already share a key $k \in \{0,1\}^n$ which is just as long as the message. Indeed, as we have seen earlier, having a key that is as long as the message is a requirement to ensure absolute security for arbitrary messages m!

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

$$\mathrm{Enc}(k,m) = m \oplus k = (m_1 \oplus k_1, m_2 \oplus k_2, \ldots, m_n \oplus k_n) = (e_1, \ldots, e_n) = e \,, \tag{1.66}$$

where $m_j \oplus k_j = m_j + k_j \mod 2$ is the XOR, or addition modulo 2. The decryption is given by

$$\mathrm{Dec}(k,e) = e \oplus k = (e_1 \oplus k_1, e_2 \oplus k_2, \ldots, e_n \oplus k_n). \tag{1.67}$$

Note that since $m_j \oplus k_j \oplus k_j = m_j$ the receiver can recover the message, thus the scheme is correct. Is it secure?

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

$$p(e|m) = p(m \oplus k|m) = p(k|m) = \frac{1}{2^n} \,, \tag{1.68}$$

for $k = e \oplus m$. Now note that this holds for all messages m, and hence

$$p(e) = \sum_m p(m)p(e|m) = \frac{1}{2^n} \,. \tag{1.69}$$

Applying Bayes rule we thus have

$$p(m|e) = \frac{p(m,e)}{p(e)} = \frac{p(e|m)p(m)}{p(e)} = p(m) \,, \tag{1.70}$$

independent of m, and since $p(m|e) = p(m)$ the scheme is perfectly secure. Note however 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 week 4 we will learn about a method called *privacy amplification* that can be used to "improve" the quality of a key about which the eavesdropper may have partial information. We will make this notion precise later in this lecture series!

We note that while the one-time pad is perfectly secure, it does not protect against the eavesdropper changing bits in the mesages. For example, Eve can flip bits - while this may not bother you very much when transmitting images, it surely will be an issue in your bank transactions. For this reason, one-time pads are supplemented by checksums or message authentication codes (MAC) which allow changes to be detected (and corrected). These are purely classical techniques, and hence we will not cover them 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. Let us explain this by considering the encryption of a single-bit message $m \in \{0,1\}$. Recall that we could encode the message into a quantum state as $|m\rangle$, or 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 $k = 1$. That is, when $k = 1$ we transform the state to $X|m\rangle$, or, as a density matrix, $X|m\rangle\langle m|X$. 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 density matrix

$$\rho = \frac{1}{2}\sum_{k\in\{0,1\}} X^k \rho X^k = \frac{1}{2}|m\rangle\langle m| + \frac{1}{2}X|m\rangle\langle m|X = \frac{\mathbb{I}}{2}\,. \tag{1.71}$$

Note that this density matrix ρ does *not* depend on m! That is, absolutely no information about m can be gained from the density matrix that represents the eavesdropper's view of the system, i.e. the message m and any information held by the eavesdropper is uncorrelated. This "uncorrelated-ness" is precisely the desired hallmark of an encryption scheme, and you will soon learn how to make this precise!

1.8.2 The quantum one-time pad

Let us consider the task of encrypting a qubit, instead of a classical bit [Amb+00; BR00]. In the videos, we saw a geometric argument for encrypting a qubit. Here, we will give a formal argument. Instead of one key bit, however, it turns out that we require two key bits k_1k_2 to encrypt a qubit. Indeed, it can be shown that two key bits are *necessary*. An intuition on why we need more than one key bit is that we wish to hide information in all possible bases the qubit could be 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 could for example attempt to encrypt a bit expressed in the Hadamard basis, 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.

The trick to a quantum one-time pad is then to apply a bit flip in both bases, standard and Hadamard. This can be achieved by applying $X^{k_1}Z^{k_2}$. When k_1k_2 is chosen uniformly at random, an arbitrary single-qubit ρ is encrypted to

$$\frac{1}{4}\sum_{k_1,k_2\in\{0,1\}} X^{k_1}Z^{k_2}\rho Z^{k_2}X^{k_1}\,. \tag{1.72}$$

To see why this works, let us recall 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

$$X + XXX + ZXZ + XZXZX = X + X - ZZX - XZZXX = X + X - X - X = 0\,, \tag{1.73}$$

where we use the fact that the Pauli matrices are observables (i.e. they are Hermitian and square to identity), and $\{X,Z\} = XZ + ZX = 0$. For some intuition, refer to Figure 1.1 for a visualization.

Exercise 1.8.1 Show that for all $M \in \{X,Z,Y\}$ we have

$$\frac{1}{4}\sum_{k_1,k_2} X^{k_1}Z^{k_2}MZ^{k_2}X^{k_1} = 0. \tag{1.74}$$

■

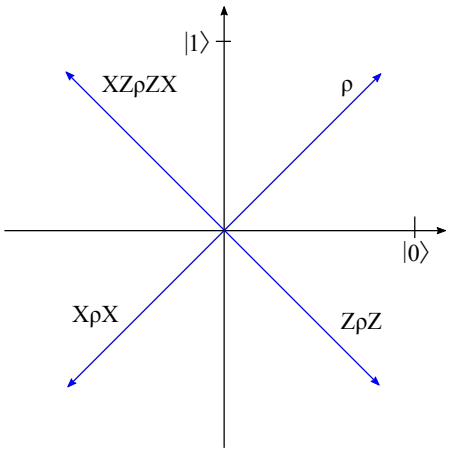

Figure 1.1: A qubit encoded by two key bits: the operations $\mathbb{I}, X, Z, XZ$ are performed on the qubit with equal probability. The resulting mixture of states is then the maximally mixed state (represented by the origin of the diagram).

Since we can write any one qubit state as

$$\rho = \frac{1}{2}\left(\mathbb{I} + v_x X + v_y Y + v_z Z\right) , \tag{1.75}$$

we thus have that

$$\frac{1}{4}\sum_{k_1,k_2} X^{k_1} Z^{k_2} \rho Z^{k_2} X^{k_1} = \frac{\mathbb{I}}{2} . \tag{1.76}$$

To someone who does not know k_1, k_2 the resulting state is again completely independent of the input ρ, which means we have managed to hide all possible information from the eavesdropper. We thus have the following encryption scheme.

Protocol 2 The quantum one-time pad is an encryption scheme for qubits. To encrypt, Alice applies $X^{k_1} Z^{k_2}$ to the qubit ρ and sends the resulting state to Bob. To decrypt, Bob applies the inverse $(X^{k_1} Z^{k_2})^\dagger$ to obtain ρ.

This scheme can be extended to n qubits, where on each qubit we apply either $\mathbb{I}$, X, Z or XZ depending on two key bits. This means that to encrypt n qubits, we use $2n$ bits of classical key.

Exercise 1.8.2 Show that strings of Pauli matrices $P^s = X^{s_1} Z^{s_2} \otimes X^{s_3} Z^{s_4} \otimes \ldots \otimes X^{s_{2n-1}} Z^{s_{2n}}$ with $s \in \{0,1\}^{2n}$ form an orthogonal basis for all linear operators $\mathscr{L}(\mathbb{C}^{2^n}, \mathbb{C}^{2^n})$, in which n-qubit density matrices ρ can be described. That is, $\mathrm{tr}[(P^s)^\dagger P^{\hat{s}}] = 0$ for all $s \neq \hat{s}$, and that we can write a density matrix on n qubits as

$$\rho = \frac{1}{2^n}\left(\mathbb{I}^{\otimes 2n} + \sum_{s \neq 0} v_s P^s\right) . \tag{1.77}$$

■

R 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.

Acknowledgements

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence. The lecture notes are written by Stephanie Wehner, Nelly Ng, and Thomas Vidick. We thank David Elkouss, Jonas Helsen, Jérémy Ribeiro and Kenneth Goodenough for proofreading.

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

$$\rho = \sum_x p_x \rho_x. \tag{1.78}$$

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

$$\rho = \frac{1}{2}\left(\mathbb{I} + v_x X + v_z Z + v_y Y\right), \tag{1.79}$$

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

$$p_{w_j} = \langle w_j|\rho|w_j\rangle = \mathrm{tr}(\rho|w_j\rangle\langle w_j|). \tag{1.80}$$

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

$$\rho_{AB} = \rho_A \otimes \rho_B := \begin{pmatrix} a_{11}\rho_B & a_{12}\rho_B \\ a_{21}\rho_B & a_{22}\rho_B \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}. \tag{1.81}$$

Partial trace

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

$$\rho_{AB} = \sum_{ijkl} \gamma_{ij}^{kl} |i\rangle\langle j| \otimes |k\rangle\langle l|, \tag{1.82}$$

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

$$\rho_B = \mathrm{tr}_A(\rho_{AB}) = \sum_{ijk\ell} \gamma_{ij}^{k\ell}\, \mathrm{tr}(|i\rangle\langle j|) \otimes |k\rangle\langle \ell| = \sum_{k\ell} \left(\sum_j \gamma_{jj}^{k\ell}\right) |k\rangle\langle \ell|\,. \tag{1.83}$$

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: $S_1^2 = \mathbb{I}$.

Bibliography

[Amb+00] A. Ambainis et al. “Private quantum channels”. In: *Proceedings of FOCS*. arXiv:quant-ph/0003101. 2000 (cited on page 20).

[BR00] P. O. Boykin and V. Roychowdhury. “Optimal encryption of quantum bits”. quant-ph/0003059. 2000 (cited on page 20).

[Kel94] D.G. Kelly. *Introduction to Probability*. Macmillan Publishing Company, 1994. ISBN: 9780023631450 (cited on page 3).

[Ros10] S.M. Ross. *A First Course in Probability*. Pearson Prentice Hall, 2010. ISBN: 9780136033134 (cited on page 3).

[Sha49] Claude E Shannon. “Communication theory of secrecy systems”. In: *Bell system technical journal* 28.4 (1949), pages 656–715 (cited on page 17).