QUANTUM CIRCUITS, COMMUNICATION & ERROR CORRECTION: PREREQUISITES

1
Probability

This chapter gives a concise review of discrete probability, covering the definitions and tools that will appear throughout our workshop. For a thorough treatment of discrete probability, please see Mitzenmacher and Upfal [1] (Chapters 1-4) or Lehman, Leighton and Meyer [2] (Chapters 17-20).

[1] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, 2nd edition, 2017

[2] Eric Lehman, F. Thomson Leighton, and Albert R. Meyer. *Mathematics for Computer Science*. MIT OpenCourseWare, 2018. Available at `https://courses.csail.mit.edu/6.042/spring18/mcs.pdf`

1.1 Sample spaces, events, and probability distributions

Probability is formally defined over a set of possible outcomes Ω of a (discrete) random experiment; Ω is called the *sample space*. For example, the outcomes of a single coin flip are heads (H) or tails (T), giving $\Omega = \{H, T\}$. If we instead flip a coin n times independently, the sample space is

$$\Omega = \{(\omega_1, \ldots, \omega_n) : \omega_i \in \{H, T\}\}.$$

Definition 1 (Probability Distribution)**.** A *probability distribution* over a finite sample space Ω is a function $p : \Omega \to \mathbb{R}_{\geq 0}$ satisfying

$$\sum_{x \in \Omega} p(x) = 1.$$

An *event* is any subset $A \subseteq \Omega$. Its probability is

$$\Pr[A] = \sum_{x \in A} p(x).$$

Typically, they represent the outcomes of a random experiment. When the distribution is *uniform* (i.e. $p(x) = 1/|\Omega|$ for every x, meaning an element from Ω is picked uniformly at random), this simplifies to $\Pr[A] = |A|/|\Omega|$.

Definition 2 (Independence of events)**.** Two events $A, B \subseteq \Omega$ are *independent* if $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$.

Terms/notations introduced:
sample space: Ω
probability distribution: p
event: $A \subseteq \Omega$
independence of events

Exercise 1. Consider tossing a fair coin three times. Let A be the event "the first flip is heads" and B the event "the total number of heads is even." Are A and B independent?

1.2 Random variables and independence

A *(real-valued) random variable* is a function $X : \Omega \to \mathbb{R}$. For example, in the experiment of tossing a fair coin n times, the number of heads is a random variable:

$$X((\omega_1, \ldots, \omega_n)) = |\{i : \omega_i = H\}|.$$

A random variable is a *function* of the outcome; it is not itself random. The randomness comes from the underlying sample space.

Definition 3 (Independence of random variables)**.** Two discrete real-valued random variables X, Y are *independent* if for all $x, y \in \mathbb{R}$,

$$\Pr[X = x,\ Y = y] = \Pr[X = x] \cdot \Pr[Y = y].$$

Random variables $X_1, \ldots, X_n$ are *jointly independent* (or *mutually independent*) if for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \prod_{i=1}^{n} \Pr[X_i = x_i].$$

They are *pairwise independent* if every pair (X_i, X_j) with $i \neq j$ is independent.

Pairwise independence does *not* imply joint independence! See Exercise 2.

Exercise 2. Let $\Omega = \{0,1\}^2$ with the uniform distribution. Define $X(\omega_1, \omega_2) = \omega_1$, $Y(\omega_1, \omega_2) = \omega_2$, and $Z(\omega_1, \omega_2) = \omega_1 \oplus \omega_2$ (XOR). Show that X, Y, Z are pairwise independent but not jointly independent.

1.3 Law of total probability and conditional probability

1.3.1 Law of total probability

If events $A_1, A_2, \ldots, A_n$ form a partition of Ω (that is, they are pairwise disjoint and their union is Ω), then for any event B,

$$\Pr[B] = \sum_{i=1}^{n} \Pr[B \cap A_i].$$

1.3.2 Conditional probability

The probability of event A *conditioned on* event B (with $\Pr[B] > 0$) is

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

Intuitively, conditioning on B restricts the sample space to the outcomes in B and normalizes the probabilities by dividing the probability of each outcome by $\Pr[B]$.

From this definition, we immediately obtain *Bayes' rule*:

$$\Pr[A \mid B] = \frac{\Pr[B \mid A]\ \Pr[A]}{\Pr[B]},$$

and the *chain rule* of conditional probability:

$$\Pr[A_1 \cap A_2 \cap \cdots \cap A_n] = \Pr[A_1]\ \Pr[A_2 \mid A_1]\ \Pr[A_3 \mid A_1 \cap A_2] \cdots \Pr[A_n \mid A_1 \cap \cdots \cap A_{n-1}].$$

Exercise 3. Suppose we have two coins: Coin 1 is fair, and Coin 2 has heads probability 2/3. We pick a coin uniformly at random and toss it twice. Given that both tosses land heads, what is the probability that we picked Coin 2?

1.4 Union bound

For any events $A_1, A_2, \ldots, A_n \subseteq \Omega$,

$$\Pr\left[\bigcup_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \Pr[A_i].$$

This is called the *union bound*. It follows from the inclusion-exclusion principle. The union bound is a simple but surprisingly powerful tool: when we want to argue that "none of the bad events $A_1, \ldots, A_n$ occurs," we can write

$$\Pr\left[\bigcap_{i=1}^{n} \overline{A_i}\right] = 1 - \Pr\left[\bigcup_{i=1}^{n} A_i\right] \geq 1 - \sum_{i=1}^{n} \Pr[A_i].$$

The union bound is tight when the events A_i are pairwise disjoint.

1.5 Expectation

Definition 4 (Expectation)**.** For a discrete real-valued random variable X taking values $x_1, \ldots, x_n$, the *expectation* (or *expected value*) of X is

$$\mathbb{E}[X] = \sum_{i=1}^{n} \Pr[X = x_i] \cdot x_i.$$

Linearity of expectation holds *regardless* of whether the random variables are independent.

The most important property of expectation is its *linearity*: for any random variables $X_1, \ldots, X_n$ (not necessarily independent) and any constants $c_1, \ldots, c_n \in \mathbb{R}$,

$$\mathbb{E}\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i \, \mathbb{E}[X_i].$$

When X and Y are *independent*, we additionally have *multiplicativity*: $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

1.5.1 Indicator random variables

Suppose A is an event. The indicator random variable for A is defined as follows:

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Indicator random variables help "arithmetize" boolean operations.

$$\begin{aligned} \mathbb{I}_{\overline{A}} &= 1 - \mathbb{I}_A; \\ \mathbb{I}_{A \cap B} &= \mathbb{I}_A \cdot \mathbb{I}_B; \\ \mathbb{I}_{A \cup B} &= 1 - (1 - \mathbb{I}_A) \cdot (1 - \mathbb{I}_B). \end{aligned}$$

Note that here 1 stands for the constant 1 random variable that maps every sample point to 1. Indicator random variables allow

Exercise 4. Suppose we toss a fair coin n times. Using linearity of expectation, show that the expected number of heads is $n/2$.

Exercise 5 (Geometric random variable)**.** Suppose we have a biased coin with $\Pr[\text{heads}] = p > 0$. Let T be the number of tosses until the first heads (including the toss that results in heads). Show that $\mathbb{E}[T] = 1/p$.

Exercise 6 (Coupon collector)**.** Suppose there are n distinct types of coupons, and at each step we receive a coupon of a uniformly random type (independently). Let T be the time until we have collected at least one coupon of every type. Using Exercise 5 and linearity of expectation, show that $\mathbb{E}[T] = nH_n$, where $H_n = \sum_{k=1}^{n} 1/k = \Theta(\log n)$ is the n-th harmonic number.

Exercise 7. (a) Show that for all events A, we have $\Pr[A] = \mathbb{E}[\mathbb{I}_A]$

(b) Show that for all events A, B, and all ω, we have

$$\begin{aligned} \mathbb{I}_{A \cup B}(\omega) &= 1 - (1 - \mathbb{I}_A(\omega))(1 - \mathbb{I}_B(\omega)) \\ &\leq \mathbb{I}_A(\omega) + \mathbb{I}_B(\omega). \end{aligned}$$

(c) Derive the union bound $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$.

(d) Generalize the above derivation and conclude

$$\begin{aligned} \Pr[\overline{A_1 \cup A_2 \cup \cdots \cup A_n}] &= \mathbb{E}[(1 - \mathbb{I}_{A_1})(1 - \mathbb{I}_{A_2}) \\ &\quad \cdots (1 - \mathbb{I}_{A_n})] \\ &= \sum_{S \subseteq [n]} (-1)^{|S|} \Pr\left[\bigcap_{i \in S} A_i\right]. \end{aligned}$$

us to use manipulations using expectations to estimate the probability of events: $\Pr[A] = \mathbb{E}[\mathbb{I}_A]$. For example, to say that the events $A_1, A_2, \ldots, A_n$ partition the probability space is equivalent to $\mathbb{I}_{A_1} + \mathbb{I}_{A_2} + \cdots + \mathbb{I}_{A_n} = 1$, from which it follows that

$$\mathbb{I}_B = \mathbb{I}_B(\mathbb{I}_{A_1} + \mathbb{I}_{A_2} + \cdots + \mathbb{I}_{A_n}) = \mathbb{I}_B\mathbb{I}_{A_1} + \mathbb{I}_B\mathbb{I}_{A_2} + \cdots + \mathbb{I}_B\mathbb{I}_{A_n}.$$

By taking expectations of the LHS and the RHS, we obtain the law of total probability.

1.6 Variance

Definition 5 (Variance)**.** The *variance* of a random variable X is

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Variance measures how spread out a random variable is around its mean. An important property: if $X_1, \ldots, X_n$ are *pairwise independent*, then

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i].$$

Variance is additive under *pairwise independence* — joint independence is not required. This is because the cross-terms $\mathbb{E}[X_iX_j]$ factor as $\mathbb{E}[X_i]\,\mathbb{E}[X_j]$ whenever X_i and X_j are independent.

Exercise 8. Let X be a Bernoulli random variable with $\Pr[X = 1] = p$ and $\Pr[X = 0] = 1 - p$. Show that $\mathrm{Var}[X] = p(1 - p)$.

Exercise 9. Let X be the uniform random variable on $\{1, 2, \ldots, n\}$. Compute $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

1.7 Concentration inequalities

A recurring theme in both classical and quantum computing is the following: a random variable that is a sum of many independent (or weakly dependent) terms is very unlikely to deviate far from its expectation. The following inequalities make this precise in increasing order of strength.

Theorem 1 (Markov's Inequality)**.** *Let Y be a non-negative random variable. Then for any $a > 0$,*

$$\Pr[Y \geq a] \leq \frac{\mathbb{E}[Y]}{a}.$$

Markov's inequality is the weakest of the three bounds we present, but it requires the least information about Y (only its expectation).

Exercise 10. Let Y be a non-negative random variable. Then, for all $a > 0$, we have $\mathbb{I}_{Y \geq a} \leq Y/a$ (with probability 1). Derive Markov's inequality from this by taking expectations. Similarly, for $k > 0$, devise a suitable upper bound for $\mathbb{I}_{|X-\mu| \geq k}$ and conclude Chebyshev's inequality from it by taking expectations.

Theorem 2 (Chebyshev's Inequality)**.** *Let X be a random variable with mean $\mu = \mathbb{E}[X]$ and variance $\sigma^2 = \mathrm{Var}[X]$. Then for any $k > 0$,*

$$\Pr[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}.$$

Chebyshev's inequality follows from applying Markov's inequality to the non-negative random variable $(X - \mu)^2$.

When the random variable is a sum of mutually independent $\{0,1\}$-valued random variables, we can obtain much tighter, exponentially decaying tail bounds.

Theorem 3 (Chernoff bounds)**.** *Let $X_1, \ldots, X_n$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[X]$. Then for any $0 < \beta < 1$,*

$$\Pr[X > (1+\beta)\mu] < e^{-\beta^2\mu/3},$$
$$\Pr[X < (1-\beta)\mu] < e^{-\beta^2\mu/2}.$$

For the upper tail with $\beta > 1$, the bound becomes $\Pr[X > (1+\beta)\mu] < e^{-\beta\mu/3}$.

Chernoff bounds require full mutual independence of the summands, unlike Chebyshev which only needs pairwise independence.

Exercise 11. Consider the coupon collector problem from Exercise 5. Let T be the time to collect all n coupon types. Show that $\mathrm{Var}[T] \leq 2n^2$, and use Chebyshev's inequality to conclude that $\Pr[|T - nH_n| \geq cn] \leq 2/c^2$ for any $c > 0$.

Exercise 12. Suppose we run a quantum algorithm n times independently, and each run succeeds with probability $2/3$. We take a majority vote over the n outcomes. Using Chernoff bounds, show that the probability of an incorrect majority answer decreases *exponentially* in n.

Exercise 13. We flip a biased coin that lands heads with probability $1/3$ a total of n times. Using Chernoff bounds, determine a value of n such that the probability of getting more than half of the flips as heads is less than $1/1000$.

2
Linear algebra

This linear algebra overview is a brief crash course tailored for our Quantum FDP workshop. For a thorough, comprehensive treatment of linear algebra, please refer to Artin [1], Chapters 3-5.

[1] Michael Artin. *Algebra*. Birkhäuser, 1998

2.1 *Complex numbers*

In quantum computing, we work extensively with complex numbers. Let us quickly recall the fundamental definitions.
A *complex number* is an expression of the form $z = a + ib$, where $a, b \in \mathbb{R}$ and the *imaginary unit* i satisfies $i^2 = -1$. We denote the set of all complex numbers by $\mathbb{C}$. The scalar $a = \mathrm{Re}(z)$ is called the *real part* of z, and $b = \mathrm{Im}(z)$ is its *imaginary part*.

In fact, one can implement any quantum algorithm using only real numbers, but doing so often makes the representation and analysis cumbersome.

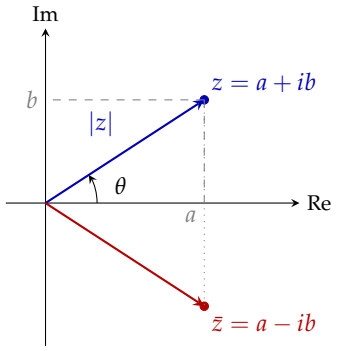

Figure 2.1: A complex number z and its complex conjugate $\bar{z}$ in the complex plane.

Definition 6 (Complex conjugate)**.** The *complex conjugate* of $z = a + ib$ is defined as

$$z^* = a - ib.$$

Geometrically, z^* represents the reflection of z across the real axis. Complex conjugation distributes over addition and multiplication ($(z_1 + z_2)^* = z_1^* + z_2^*$, $(z_1 z_2)^* = z_1^* z_2^*$), satisfies $((z)^*)^* = z$, and allows us to isolate components via $z + z^* = 2\,\mathrm{Re}(z)$ and $z - z^* = 2i\,\mathrm{Im}(z)$.

Definition 7 (Absolute value)**.** The *absolute value* (or *modulus*) of $z = a + ib$ is the non-negative real number

$$|z| = \sqrt{a^2 + b^2},$$

or equivalently, $|z|^2 = z \cdot z^*$.

Geometrically, $|z|$ corresponds to the Euclidean distance from z to the origin. The modulus is multiplicative ($|z_1 z_2| = |z_1||z_2|$) and obeys the triangle inequality: $|z_1 + z_2| \leq |z_1| + |z_2|$.

Exercise 14. Find complex numbers $z_1, z_2 \in \mathbb{C}$ such that the (strict) inequality $|z_1 + z_2| < |z_1| + |z_2|$ holds; under what condition does $|z_1 + z_2| = |z_1| + |z_2|$ hold?

Definition 8 (Polar and exponential forms)**.** Every complex number $z \neq 0$ can be uniquely expressed in terms of its magnitude $r = |z|$

and an angle $\theta \in [0, 2\pi)$ called the *argument* (denoted $\arg(z)$):

$$z = r(\cos\theta + i\sin\theta) = re^{i\theta}.$$

Here, the identity $e^{i\theta} = \cos\theta + i\sin\theta$ is Euler's formula.

2.2 *Vectors, matrices and subspaces*

2.2.1 *Vectors*

In quantum computing, the state of a system is represented by a vector with complex entries. Formally, an n-dimensional vector is an ordered collection of n complex numbers, arranged in a single vertical column:

$$v = \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_{n-1} \end{pmatrix},$$

where each component $v_i \in \mathbb{C}$. The set of all such vectors is the n-dimensional complex space, denoted by $\mathbb{C}^n$.

We will use $|\cdot\rangle$ and $\langle\cdot|$ to keep track of column vectors and row vectors respectively.

We denote a column vector using *ket notation*, written as $|v\rangle$. The conjugate transpose (or adjoint) of $|v\rangle$ is a row vector whose i^{th} component is ${v_i}^*$. We use the *bra notation* to denote the conjugate transpose of $|v\rangle$, that is, $\langle v| = |v\rangle^\dagger$. When dealing with qubits, $\langle 0|$ and $|0\rangle$ have a different meaning (in fact, they are unit vectors); to avoid confusion, we will write $\mathbf{0}^\dagger$ and $\mathbf{0}$ for the all-zeroes row vector and the all-zeroes column vector.

Vectors of the same dimension can be added together via vector (component-wise) addition, and a vector can be scaled by any complex number (also called a *scalar*) via scalar multiplication. For example, consider two vectors $|u\rangle, |v\rangle \in \mathbb{C}^n$ and scalars $\alpha, \beta \in \mathbb{C}$. Then

$$\alpha|u\rangle + \beta|v\rangle = \begin{pmatrix} \alpha u_0 + \beta v_0 \\ \alpha u_1 + \beta v_1 \\ \vdots \\ \alpha u_{n-1} + \beta v_{n-1} \end{pmatrix}.$$

If $|v_1\rangle, |v_2\rangle, \ldots, |v_k\rangle$ is linearly independent, then the *only* way to obtain the zero vector as their linear combination is to multiply each by the scalar 0 and adding them up.

Definition 9 (Linear independence)**.** A sequence of vectors $(|v_1\rangle, |v_2\rangle, \ldots, |v_k\rangle)$ (where each $|v_i\rangle \in \mathbb{C}^n$) is said to be *linearly independent* if for all $(c_1, c_2, \ldots, c_k) \in \mathbb{C}^k$ such that $(c_1, c_2, \ldots, c_k) \neq (0, 0, \ldots, 0)$,

$$c_1|v_1\rangle + c_2|v_2\rangle + \cdots + c_k|v_k\rangle \neq \mathbf{0}.$$

Suppose the sequence of vectors $(|v_1\rangle, |v_2\rangle, \ldots, |v_k\rangle)$ is linearly independent. Then (by definition) the *only* way to obtain the zero vector $\mathbf{0}$ as a linear combination is to multiply each by the scalar 0 and add them up.

Exercise 15. Show that if

$$\begin{aligned} &c_1|v_1\rangle + c_2|v_2\rangle + \cdots + c_k|v_k\rangle \\ =&d_1|v_1\rangle + d_2|v_2\rangle + \cdots + d_k|v_k\rangle, \end{aligned}$$

then $(c_1, c_2, \ldots, c_k) = (d_1, d_2, \ldots, d_k)$. That is, a vector can be written as a linear combination of the $|v_i\rangle$'s in at most one way.

Exercise 16. Let $n = 2^k$ for some integer $k \geq 1$. Consider the n vectors $|v_0\rangle, |v_1\rangle, \ldots, |v_{n-1}\rangle \in \mathbb{C}^n$, where the j-th component of the vector $|v_i\rangle$ is defined as:

$$v_{i,j} = (-1)^{i \cdot j \bmod 2} \quad \text{for } i, j \in \{0, 1, \ldots, n-1\}.$$

Here, $i \cdot j$ denotes the bit-wise dot product of the binary representations of i and j. Prove that these n vectors are linearly independent. (Note that as i ranges over $\{0, 1, \ldots, n-1\}$, its binary representation ranges over all n-tuples in $\{0,1\}^k$.)

This matrix is known as the Hadamard matrix H_{2^k}.

2.2.2 *Matrices*

Matrices are used to represent quantum operations. A matrix $\mathsf{A} \in \mathbb{C}^{m \times n}$ acts as a map that takes vectors from $\mathbb{C}^n$ to $\mathbb{C}^m$. The action of a matrix A on a vector $|v\rangle$ is simply written as $\mathsf{A}|v\rangle$, which produces another column vector (a new ket). The *adjoint* of a matrix A, denoted $\mathsf{A}^\dagger$, is formed by taking the transpose and then the complex conjugate of each element. The adjoint operation behaves as follows with respect to matrix addition and multiplication:

Terms/notations introduced:
adjoint of A : $\mathsf{A}^\dagger$,
inverse of A : A^{-1}

- It distributes directly over addition: $(\mathsf{A} + \mathsf{B})^\dagger = \mathsf{A}^\dagger + \mathsf{B}^\dagger$.
- It reverses the order of multiplication: $(\mathsf{AB})^\dagger = \mathsf{B}^\dagger \mathsf{A}^\dagger$.

The *identity matrix* I_n is the $n \times n$ matrix with 1s on the diagonal and 0s elsewhere; it satisfies $\mathsf{AI} = \mathsf{IA} = \mathsf{A}$ for every $\mathsf{A} \in \mathbb{C}^{n \times n}$ (we drop the subscript when the dimension is clear). A square matrix A is *invertible* if there exists a matrix A^{-1} such that $\mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = \mathsf{I}$. When it exists, this inverse is unique.

2.3 *Inner product, outer product and tensor product*

2.3.1 *Inner product*

We discussed earlier that quantum states are represented as complex vectors. Given two states $|u\rangle$ and $|v\rangle$, a natural question is how "similar" are the two states (we will make this more formal later on). This is captured by the *inner product* of the two vectors.

Definition 10 (Inner product)**.** The *inner product* of two vectors $|u\rangle, |v\rangle \in \mathbb{C}^n$ is a complex scalar:

$$\langle u|v\rangle = \sum_{i=1}^{n} u_i^* \cdot v_i.$$

Terms/notations introduced:
inner product: $\langle u|v\rangle$
norm of $|v\rangle$: $\||v\rangle\|$
orthogonal vectors

The inner product satisfies $\langle u|v\rangle = \langle v|u\rangle^*$. It is linear in the second argument: if $|v\rangle = \alpha|v_1\rangle + \beta|v_2\rangle$, then $\langle u|v\rangle = \alpha\langle u|v_1\rangle + \beta\langle u|v_2\rangle$; it

is conjugate-linear in the first argument: if $|u\rangle = \alpha|u_1\rangle + \beta|u_2\rangle$, then $\langle u|v\rangle = \alpha^*\langle u_1|v\rangle + \beta^*\langle u_2|v\rangle$. Two vectors are *orthogonal* if $\langle u|v\rangle = 0$. The *norm* (or length) of a vector is defined as $\||v\rangle\| = \sqrt{\langle v|v\rangle}$; note that the norm of a vector is non-negative; it is 0 iff the vector is the zero vector $\mathbf{0}$. All quantum states have norm 1.

Exercise 17. Suppose $|u\rangle$ is a unit vector, i.e., $\||u\rangle\|^2 = 1$. The vector $|v^{\|}\rangle = \langle u|v\rangle|u\rangle$ is said to be the *component of* $|v\rangle$ *along* $|u\rangle$, and $|v^{\perp}\rangle = |v\rangle - |v^{\|}\rangle$ the *component of* $|v\rangle$ *orthogonal to* $|u\rangle$. Verify the following:

$$\langle v^{\|}|v^{\perp}\rangle = 0;$$
$$\||v\rangle\|^2 = \||v^{\|}\rangle\|^2 + \||v^{\perp}\rangle\|^2.$$

Conclude that for any vectors $|u\rangle$ and $|v\rangle$ we have

$$|\langle u|v\rangle| \leq \||u\rangle\| \cdot \||v\rangle\|\,; \qquad \text{(Cauchy-Schwarz inequality)}$$
$$\||u\rangle + |v\rangle\| \leq \||u\rangle\| + \||v\rangle\|\,. \qquad \text{(Triangle inequality)}$$

To derive the first inequality, it might be helpful to first normalize u.

2.3.2 *Outer product*

Another related quantity is the *outer product*. We will discuss the quantum relevance a bit later in this document.

Definition 11 (Outer product)**.** The *outer product* of $|u\rangle \in \mathbb{C}^m$ and $|v\rangle \in \mathbb{C}^n$ forms an $m \times n$ matrix, written as $|u\rangle\langle v|$.

The outer product acts as a linear operator. If we apply the matrix $|u\rangle\langle v|$ to a vector $|w\rangle$, then via associativity, we get that $|u\rangle\langle v||w\rangle = |u\rangle(\langle v|w\rangle)$, which is the vector $|u\rangle$ scaled by the complex number $\langle v|w\rangle$.

2.3.3 *Tensor product*

Suppose we have one quantum system $\mathcal{A}$ whose state is described by the vector $|v\rangle$, and a separate quantum system $\mathcal{B}$ whose state is described by the vector $|w\rangle$. A natural question to ask is the following: what is the *joint* state of the two systems? Formally, this is given by the tensor product of $|v\rangle$ and $|w\rangle$, which is defined below.

We will formally discuss what it means for two systems to be separate/independent.

Definition 12 (Tensor product of vectors)**.** Given $|u\rangle \in \mathbb{C}^m$ and $|v\rangle \in \mathbb{C}^n$, their *tensor product* (or *Kronecker product*) $|u\rangle \otimes |v\rangle \in \mathbb{C}^{mn}$ is the vector obtained by multiplying each entry of $|u\rangle$ by the entire vector $|v\rangle$:

$$|u\rangle \otimes |v\rangle \;=\; \begin{pmatrix} u_1\, v \\ u_2\, v \\ \vdots \\ u_m\, v \end{pmatrix}.$$

We similarly define the tensor product of bra vectors.

$$\langle u| \otimes \langle v| \;=\; (u_1^* \, v^\dagger, \, u_2 \, v^\dagger, \, \ldots, \, u_m \, v^\dagger).$$

Earlier, we had discussed that quantum operations are represented by matrices. Continuing with our independent quantum systems, suppose an operation A acts on system $\mathcal{A}$ and an operation B acts of $\mathcal{B}$, then what is the 'combined' operation on the two systems? This is given by the tensor product of A and B, which is defined below.

Definition 13 (Tensor product of matrices)**.** Given an $m \times n$ matrix A and a $p \times q$ matrix B, their tensor product $\mathsf{A} \otimes \mathsf{B}$ is the $mp \times nq$ block matrix obtained by replacing each entry a_{ij} of A with the $p \times q$ block $a_{ij}\,\mathsf{B}$:

$$\mathsf{A} \otimes \mathsf{B} \;=\; \begin{pmatrix} a_{11}\,\mathsf{B} & a_{12}\,\mathsf{B} & \cdots & a_{1n}\,\mathsf{B} \\ a_{21}\,\mathsf{B} & a_{22}\,\mathsf{B} & \cdots & a_{2n}\,\mathsf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\,\mathsf{B} & a_{m2}\,\mathsf{B} & \cdots & a_{mn}\,\mathsf{B} \end{pmatrix}.$$

Below, we state some useful properties of the tensor product.

Theorem 4 (Bilinearity)**.** *The tensor product is linear in its first and second arguments. For vectors $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^m, |\phi_1\rangle, |\phi_2\rangle \in \mathbb{C}^n$ and scalars $c, c_1, c_2 \in \mathbb{C}$, we have*

$$|\psi_1\rangle \otimes (c_1|\phi_1\rangle + c_2|\phi_2\rangle) = c_1|\psi_1\rangle \otimes |\phi_1\rangle + c_2|\psi\rangle \otimes |\phi_2\rangle\,,$$
$$(c_1|\psi_1\rangle + c_2|\psi_2\rangle) \otimes \phi_1 = c_1|\psi_1\rangle \otimes |\phi_1\rangle + c_2|\psi_2\rangle \otimes |\phi_1\rangle\,.$$

Theorem 5 (The mixed-Product property & adjoint distribution)**.** *If* A, C *are operators acting on a vector space $\mathcal{V}$, and* B, D *are operators acting on a vector space $\mathcal{W}$, then their parallel compositions satisfy the mixed-product property:*

$$(\mathsf{A} \otimes \mathsf{B})(\mathsf{C} \otimes \mathsf{D}) = (\mathsf{A} \cdot \mathsf{C}) \otimes (\mathsf{B} \cdot \mathsf{D})$$

Furthermore, the conjugate transpose distributes directly across a tensor product without reversing the order of the constituent spaces:

$$(\mathsf{A} \otimes \mathsf{B})^\dagger = \mathsf{A}^\dagger \otimes \mathsf{B}^\dagger$$

2.4 Eigenvalues and eigenvectors

Next, we recall the notions of eigenvectors and eigenvalues.

Definition 14 (Eigenvalue, eigenvector, trace)**.** For a square matrix $\mathsf{A} = (a_{ij} : i, j = 1, 2, \ldots, n) \in \mathbb{C}^{n \times n}$, a non-zero vector $|v\rangle$ is an

Exercise 18. Let $\langle v| = (1, 0, 0, 1)$. Prove that $|v\rangle$ cannot be expressed as $|x\rangle \otimes |y\rangle$ for any $|x\rangle, |y\rangle \in \mathbb{C}^2$.

Exercise 19. Let $|v_1\rangle, |v_2\rangle, \ldots, |v_m\rangle$ be a set of m linearly independent vectors in $\mathbb{C}^m$, and similarly let $|w_1\rangle, |w_2\rangle, \ldots, |w_n\rangle$ be a set of n linearly independent vectors in $\mathbb{C}^n$. Prove that the set $\{|v_i\rangle \otimes |w_j\rangle\}_{i \in [m], j \in [n]}$ of mn vectors is linearly independent.

Exercise 20. Let $n = 2^k$, and let H_n denote the Hadamard matrix of dimension n. Prove that $\mathsf{H}_n = \underbrace{\mathsf{H}_2 \otimes \mathsf{H}_2 \otimes \cdots \otimes \mathsf{H}_2}_{k \text{ factors}}$.

Exercise 21. Show that $\||u\rangle \otimes |v\rangle\| = \||u\rangle\| \cdot \||v\rangle\|$.

Exercise 22. Consider the following derivation for scalars, $x, y, z \in \mathbb{C}$ (using the triangle inequality for complex numbers):

$$\begin{aligned} &|x-y| \cdot |z| + |y-z| \cdot |x| = |(x-y)z| + |(y-z)x| \\ &\geq |(x-y)z + (y-z)x| = |(x-z)y| \\ &= |x-z| \cdot |y|. \end{aligned}$$

Generalize this to vectors $|x\rangle, |y\rangle, |z\rangle \in \mathbb{C}^n$ and derive Ptolemy's inequality?

$$\begin{aligned} \||x\rangle - |z\rangle\| \cdot \||y\rangle\| \leq \||x\rangle - |y\rangle\| \cdot \||z\rangle\| \\ + \||y\rangle - |z\rangle\| \cdot \||x\rangle\|. \end{aligned}$$

eigenvector of A if the action of A on $|v\rangle$ only scales the vector by a complex number λ:

$$\mathsf{A}|v\rangle = \lambda|v\rangle.$$

The scalar λ is called the *eigenvalue* corresponding to $|v\rangle$. The characteristic polynomial $p_{\mathsf{A}}(x) = \det(x\,\mathbb{I}_n - \mathsf{A})$ of A can be written uniquely as $\prod_\lambda (\lambda - x)^{e_\lambda}$, where λ ranges over the set of complex eigenvalues of A. We refer to n_λ as the (algebraic) multiplicity of λ. Clearly, the number of eigen values of A counted with multiplicities is n: $\sum_\lambda n_\lambda = n$. The sum of these eigenvalues (again counted with multiplicities) is equal to the sum of the diagonal entries of A, and is referred to as the *trace of* A: $\mathsf{Tr}(\mathsf{A}) = \sum_i a_{ii} = \sum_\lambda n_\lambda \lambda$.

Exercise 23. Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be a matrix such that $\mathsf{A} = \mathsf{A}^\dagger$. A matrix that satisfies $\mathsf{A} = \mathsf{A}^\dagger$ is called a *Hermitian matrix*. Prove that all eigenvalues of a Hermitian matrix A are real.

2.5 Subspaces

A *subspace* $S \subseteq \mathbb{C}^n$ is a subset of vectors that contains $\mathbf{0}$ and is closed under addition and scalar multiplication. That is, for any $|u\rangle, |v\rangle \in S$ and any scalars $\alpha, \beta \in \mathbb{C}$, $\alpha|u\rangle + \beta|v\rangle \in S$.

Exercise 24. Let $\{|v_1\rangle, |v_2\rangle, \ldots, |v_k\rangle\}$ be k vectors in $\mathbb{C}^n$. The span of these vectors is defined as follows:

$$\mathsf{span}(|v_1\rangle, \ldots, |v_k\rangle) = \left\{ \sum_{i=1}^{k} c_i |v_i\rangle : c_i \in \mathbb{C} \right\}$$

Show that $\mathsf{span}(|v_1\rangle, \ldots, |v_k\rangle)$ is a subspace of $\mathbb{C}^n$.

Terms/notations introduced: dimension and basis of subspace span of a set of vectors

The above exercise shows one way of generating subspaces of $\mathbb{C}^n$. In fact, every subspace arises in this way: for any subspace $S \subseteq \mathbb{C}^n$, there exists a collection of linearly independent vectors $\{|b_1\rangle, \ldots, |b_k\rangle\}$ such that $S = \mathsf{span}(|b_1\rangle, \ldots, |b_k\rangle)$. Such a collection is called a *basis* of S. While the basis itself is not unique, the number of vectors in a basis is; this number is called the *dimension* of S.

Exercise 25. Let $\mathbb{C}^n$ be the standard complex vector space for $n \geq 2$.

1. Consider the subset $S \subset \mathbb{C}^n$ consisting of all vectors whose final component is zero:

$$S = \left\{ \begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \\ 0 \end{pmatrix} \;\middle|\; v_1, \ldots, v_{n-1} \in \mathbb{C} \right\}.$$

   Prove that S is a subspace of $\mathbb{C}^n$. What is the dimension of S?

2. Now consider the subset $T \subset \mathbb{C}^n$ consisting of all vectors that contain a zero in *at least one* position. Is T a subspace of $\mathbb{C}^n$?

2.5.1 Orthonormal basis of a subspace

First, let us recall the notion of an orthonormal set of vectors.

Definition 15 (Orthonormal set)**.** A set of vectors $\{|u_1\rangle, |u_2\rangle, \ldots, |u_k\rangle\}$ in $\mathbb{C}^n$ is called an *orthonormal set* if the inner product of any two vectors satisfies

$$\langle u_i | u_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Given any set of k linearly independent vectors $\{|v_1\rangle, \ldots, |v_k\rangle\}$, there exists an orthonormal set $\{|u_1\rangle, \ldots, |u_k\rangle\}$ such that $\mathsf{span}\,(|v_1\rangle, \ldots, |v_k\rangle) = \mathsf{span}\,(|u_1\rangle, \ldots, |u_k\rangle)$. From here, it follows that for any k-dimensional subspace S, there exists an orthonormal set $\{|u_1\rangle, \ldots, |u_k\rangle\}$ such that $S = \mathsf{span}\,(|u_1\rangle, \ldots, |u_k\rangle)$. The set $\{|u_1\rangle, \ldots, |u_k\rangle\}$ forms an orthonormal basis of S.

Exercise 26. Suppose $S = \{|u_1\rangle, |u_2\rangle, \ldots, |u_k\rangle\}$ is an orthonormal set. Show that S is linearly independent.

Exercise 27. Consider the $n = 2^k$ vectors $|v_1\rangle, |v_2\rangle, \ldots, |v_n\rangle$ defined in Exercise 16. Show that $\left\{ \frac{1}{\sqrt{n}}|v_1\rangle, \frac{1}{\sqrt{n}}|v_2\rangle, \ldots, \frac{1}{\sqrt{n}}|v_n\rangle \right\}$ is an orthonormal set.

2.5.2 Coordinates

Suppose $\mathsf{B} = \{|u_1\rangle, \ldots, |u_k\rangle\}$ is an orthonormal basis for a subspace S. Then, each vector $|w\rangle \in S$ can be written in a unique way in the form

$$|w\rangle = \sum_i c_i |u_i\rangle,$$

where $c_i \in \mathbb{C}$. Then, the column vector $(c_1, c_2, \ldots, c_k)^T$ gives us the coordinates of $|w\rangle$ with respect to the (ordered) basis B; we may write this as

$$|w\rangle = \mathsf{B} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = (|u_1\rangle, |u_2\rangle, \ldots, |u_k\rangle) \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$$

The bra vector $\langle u_i|$ extracts the coordinate c_i from the vector $|w\rangle$.

Exercise 28. Suppose $c_1, c_2, \ldots, c_k$ are the coordinates of a vector $|w\rangle$ wrt an orthonormal basis $\mathsf{B} = \{|u_1\rangle, \ldots, |u_k\rangle\}$. Show that

$$\||w\rangle\|^2 = c_1^2 + c_2^2 + \cdots + c_k^2.$$

Exercise 29. When working with orthonormal bases determining the coordinates is particularly convenient. Show that $c_i = \langle u_i | w \rangle$. Note that the coordinate corresponding to $|u_i\rangle$ does not depend on the other vectors in the orthonormal basis. Derive the Cauchy-Schwarz inequality from this using the fact that for every unit vector $|u\rangle$, there is an orthonormal basis that contains $|u\rangle$.

2.6 Some special matrices

In this section, we will see some properties of special families of matrices.

2.6.1 Unitary matrices

Unitary matrices are at the heart of quantum computation. There are several (equivalent) definitions of unitary matrices.

Definition 16 (Unitary matrix)**.** A matrix $\mathsf{U} \in \mathbb{C}^{n \times n}$ is unitary if

$$\mathsf{U} \cdot \mathsf{U}^\dagger = \mathsf{U}^\dagger \cdot \mathsf{U} = \mathsf{I}_n.$$

Below, we list some useful properties of unitary matrices.

Exercise 30 (Unitary operation preserves the norm and inner product)**.** Let $\mathsf{U} \in \mathbb{C}^{n\times n}$ be a unitary matrix. Consider any vectors $|\phi\rangle, |\psi\rangle \in \mathbb{C}^n$, and let $|u\rangle = \mathsf{U}|\phi\rangle$, $|v\rangle = \mathsf{U}|\psi\rangle$. Show that $\langle\phi|\psi\rangle = \langle u|v\rangle$.

Recall, quantum states are vectors with norm 1. This exercise shows that unitary matrices map quantum states to quantum states. In fact, we can define unitary matrices using this property.

From the above exercise, it also follows that all eigenvalues of U have absolute value 1. In fact, one can show a stronger *spectral decomposition* property for unitary matrices.

Theorem 6. *For any unitary matrix $\mathsf{U} \in \mathbb{C}^{n\times n}$, there exist n complex numbers $\lambda_1, \ldots, \lambda_n$ and n orthonormal vectors $|u_1\rangle, \ldots, |u_n\rangle$ such that for all i, $|\lambda_i| = 1$, and*

$$\mathsf{U} = \sum_{i=1}^{n} \lambda_i |u_i\rangle\langle u_i|.$$

Exercise 31. A matrix M with entries of 0, 1 of order $n \times n$ is said to be a *permutation matrix* if for each row $i \in [n]$, there is a exactly one entry that is 1 and no two distinct rows $i, j \in [n]$, have a 1 at the same column index. Show that any such M is unitary.

Exercise 32. Recall the n vectors $|v_1\rangle, \ldots, |v_n\rangle \in \mathbb{C}^n$ defined in Exercise 16. Now, consider a matrix H where the i^{th} column of this matrix is $\frac{1}{\sqrt{n}}|v_i\rangle$ (and $n = 2^k$ for some k). Prove that H is a unitary matrix.

This is a special matrix called the Hadamard matrix, which we will encounter several times in this course. It is easy to check that $\mathsf{H} = \mathsf{H}^\top$.

Exercise 33. Let $S \subseteq \mathbb{C}^n$ be a k-dimensional subspace. Let $\{|u_1\rangle, \ldots, |u_k\rangle\}$ and $\{|v_1\rangle, \ldots, |v_k\rangle\}$ be two orthonormal bases for S. Show that there exists a unitary matrix $\mathsf{U} \in \mathbb{C}^{n\times n}$ such that for all $i \in [k]$, $\mathsf{U}|u_i\rangle = |v_i\rangle$.

This exercise shows that even though a subspace S can have different orthonormal bases, one can be obtained from another via a unitary operation.

2.6.2 Hermitian matrices

We encountered Hermitian matrices in Exercise 23 ($\mathsf{A} = \mathsf{A}^\dagger$), and saw that all eigenvalues of A are real-valued. Below, we present a spectral decomposition property similar to Theorem 6.

Theorem 7. *For any Hermitian matrix $\mathsf{H} \in \mathbb{C}^{n\times n}$, there exist n real numbers $\lambda_1, \ldots, \lambda_n$ and n orthonormal vectors $|u_1\rangle, \ldots, |u_n\rangle$ such that*

$$\mathsf{U} = \sum_{i=1}^{n} \lambda_i |u_i\rangle\langle u_i|.$$

Theorems 6 and 7 are special cases of the general *spectral theorem* (see Artin[2], Theorem 8.6.6), which states that if a matrix A satisfies $\mathsf{A}\mathsf{A}^\dagger = \mathsf{A}^\dagger\mathsf{A}$, then it can be expressed as $\mathsf{A} = \sum_i \lambda_i |v_i\rangle\langle v_i|$ where $\{|v_1\rangle, \ldots, |v_n\rangle\}$ form an orthonormal set. For the case of unitary matrices, Exercise 30 implies that $|\lambda_i| = 1$ for all i. Similarly for Hermitian matrices, we showed in Exercise 23 that $\lambda_i \in \mathbb{R}$ for all i.

[2] Michael Artin. *Algebra*. Birkhäuser, 1998

2.6.3 *Density matrices*

Density matrices are a special case of Hermitian matrices where all eigenvalues are non-negative, and their sum is equal to 1 (or equivalently, the diagonal entries of the Hermitian matrix are non-negative and add up to 1). We will encounter density matrices when we discuss mixed quantum states.

2.6.4 *Projection matrices*

A projection matrix P is a matrix that satisfies $\mathsf{P}^2 = \mathsf{P}$. As the name suggests, the projection matrix *projects* onto a subspace. For all vectors in the subspace, this matrix acts as identity. If P additionally satisfies $\mathsf{P} = \mathsf{P}^\dagger$, then we call it an *orthogonal projection* matrix.
We will mostly deal with orthogonal projection matrices. Given any such matrix P, the projection subspace is $S_\mathsf{P} = \{\mathsf{P}|v\rangle : |v\rangle \in \mathbb{C}^n\}$.
The reverse construction also works: consider a subspace S with an orthonormal basis set $\{|v_1\rangle, \ldots, |v_k\rangle\}$. Then the orthogonal projection

$$\mathsf{P}_S = \sum_{i=1}^{k} |v_i\rangle\langle v_i|$$

projects any vector onto the subspace S.

Exercise 34. A subspace can have multiple orthonormal bases. Does the corresponding orthogonal projection matrix depend on the choice of the orthonormal basis?

3
Probability through the linear algebra lens

In the previous chapters, we reviewed probability and linear algebra separately. In this chapter, we show how classical probability can be cast entirely in the language of vectors and matrices. This perspective is not merely an intellectual exercise: it serves as the bridge to quantum mechanics, where the same linear-algebraic framework is used, but with complex amplitudes replacing real probabilities.

3.1 Probability distributions as vectors

Consider a finite sample space $\Omega = \{1, 2, \ldots, n\}$. A probability distribution p over Ω assigns a non-negative real number $p(i)$ to each outcome i, with the constraint that $\sum_i p(i) = 1$. We can represent this distribution as a column vector in $\mathbb{R}^n$:

We will use $|\cdot\rangle\!\rangle$ to remind ourselves that this is a column vectors whose entries are non-negative real numbers that add up to 1. Similarly, we will use $\langle\!\langle\cdot|$ for row vectors.

$$|p\rangle\!\rangle = \begin{pmatrix} p(1) \\ p(2) \\ \vdots \\ p(n) \end{pmatrix}, \qquad \text{where } p(i) \geq 0 \text{ and } \sum_{i=1}^{n} p(i) = 1.$$

Let $|e_1\rangle\!\rangle, |e_2\rangle\!\rangle, \ldots, |e_n\rangle\!\rangle$ denote the standard basis vectors of $\mathbb{R}^n$ (with a 1 in the i-th position and 0s elsewhere). A *deterministic state* — one in which outcome i occurs with certainty — corresponds to the basis vector $|e_i\rangle\!\rangle$. Every probability distribution can then be written as a convex combination of deterministic states:

$$|p\rangle\!\rangle = \sum_{i=1}^{n} p(i)\, |e_i\rangle\!\rangle.$$

3.2 Joint systems and independence

Suppose we have two independent classical systems: system $\mathcal{A}$ with sample space $\{1, \ldots, m\}$ and distribution $|p\rangle\!\rangle$, and system $\mathcal{B}$ with sample space $\{1, \ldots, n\}$ and distribution $|q\rangle\!\rangle$. The joint distribution of

the two systems is the tensor product:

$$|p\rangle\rangle \otimes |q\rangle\rangle \in \mathbb{R}^{mn}.$$

The (i,j)-th entry of this vector is $p(i) \cdot q(j) = \Pr[\mathcal{A} = i \text{ and } \mathcal{B} = j]$. This is precisely the definition of independence: the joint probability factors as a product.

Not every joint distribution is a tensor product. A joint distribution that cannot be written as $|p\rangle\rangle \otimes |q\rangle\rangle$ describes *correlated* classical systems.

The quantum analogue of this phenomenon is *entanglement*

Exercise 35. Let $|r\rangle\rangle = \frac{1}{2}(1,0,0,1)^\top \in \mathbb{R}^4$ be a joint distribution over $\{1,2\} \times \{1,2\}$. Show that there do not exist distributions $|p\rangle\rangle \in \mathbb{R}^2$ and $|q\rangle\rangle \in \mathbb{R}^2$ such that $|r\rangle\rangle = |p\rangle\rangle \otimes |q\rangle\rangle$.

3.3 *Events, random variables and expectation*

Events: Recall, an event E is a subset of Ω, and the probability of E with respect to probability distribution p is $p_E = \sum_{a \in E} p(a)$. We can express this in our linear-algebraic framework as follows: consider the projective matrix

$$\mathsf{P}_E = \sum_{a \in E} |a\rangle\rangle\langle\langle a|.$$

Applying P_E to $|p\rangle\rangle$ results in a vector that is 0 at every position $x \notin E$, and is $p(x)$ at position $x \in E$. Therefore, note that

$$p_E = |\mathsf{P}_E|p\rangle\rangle|.$$

Random variables and expectation: Recall, random variables assign values to the outcomes of a probabilistic experiment. Given a random variable $X : \Omega \to \mathbb{R}$, the expected value of X with respect to a probability distribution p is given by $\mathbb{E}_p[X] = \sum_{a \in \Omega} X(a)p(a)$. Let $\langle\langle X|$ denote the row vector corresponding to X (that is, the entry at position a is $X(a)$). Then $\mathbb{E}_p[X] = \langle\langle X|p\rangle\rangle$. Note that with this association, linear of expectation follows immediately from the linearity in the first argument of the (real) inner product.

4
Number theory

Shor's algorithm for factoring is one of the most significant breakthroughs in quantum computing. The algorithm consists of two components: a classical reduction from the factoring problem to the *order-finding* problem, and a quantum algorithm that solves the order-finding problem efficiently. In this chapter, we develop the number-theoretic tools needed to understand this reduction.

4.1 Divisibility and greatest common divisors

We begin by recalling divisibility and the greatest common divisor. Let $\mathbb{Z}$ denote the set of integers.

Terms/notations introduced:
$a \mid b$: a divides b
$a \equiv b \pmod{N}$: N divides $(a-b)$

Definition 17 (Divisibility and Greatest Common Divisor)**.** Let $a, b \in \mathbb{Z}$. We say that a divides b, denoted as $a \mid b$, if there exists an integer k such that $b = ak$. For integers a, b (not both zero), the greatest common divisor $\gcd(a, b)$ is the largest integer d such that $d \mid a$ and $d \mid b$.

The naive approach to computing $\gcd(a, b)$ requires $O(\min(a, b))$ operations, which scales exponentially with the number of bits needed to represent the inputs. To compute the greatest common divisor efficiently, we instead use Euclid's algorithm, which runs in polynomial time in the input bit-length.

4.2 The order finding problem

Definition 18 (Congruence)**.** We write $a \equiv b \pmod{N}$ if $N \mid (a-b)$.

We now introduce the order finding problem.

Definition 19 (Order finding problem)**.** Let $a, N \in \mathbb{N}$ with $a < N$ and $\gcd(a, N) = 1$. The order finding problem is to determine the smallest positive integer r such that

$$a^r \equiv 1 \pmod{N}.$$

For any $a \in [1, N]$ such that $\gcd(a, N) = 1$, the order of a modulo N is the smallest positive r such that $a^r \equiv 1 \pmod{N}$. Such an r is guaranteed to exist (see Euler's theorem).

4.3 Reduction: factoring to order finding

Suppose we wish to factor a composite number N which is a product of two primes. We select an integer a uniformly at random from the range $1 < a < N$. We first compute $\gcd(a, N)$. If this is strictly greater than 1, we have trivially found a factor and the protocol terminates. If $\gcd(a, N) = 1$, we use the quantum order-finding subroutine to compute the order r of a modulo N. **If r is even and N does not divide $a^{r/2} + 1$**, then we are done! (If r is odd or N divides $a^{r/2} + 1$, choose a fresh random a and repeat.)

Here, we restrict ourselves to N which is a product of two primes, since this setting suffices for breaking most existing cryptosystems. However, the ideas discussed here can be extended to factor any composite number.

- Since r is the order of a modulo N, we know that N divides $a^r - 1$, but N does not divide $a^k - 1$ for any $k < r$.
- Because r is even, we can express $a^r - 1$ as $\left(a^{r/2} - 1\right)\left(a^{r/2} + 1\right)$.
- By the definition of the order r, we know N cannot divide $a^{r/2} - 1$.
- Based on our established condition, N does not divide $a^{r/2} + 1$.

Since N divides the product $\left(a^{r/2} - 1\right)\left(a^{r/2} + 1\right)$ but neither of the terms individually, it must be that some factors of N divide $a^{r/2} - 1$, while the remaining factors divide $a^{r/2} + 1$. Consequently, computing $\gcd\left(N, a^{r/2} - 1\right)$ and $\gcd\left(N, a^{r/2} + 1\right)$ will yield non-trivial factors of N.

It remains to be shown that with a sufficiently high probability, if we choose an integer $a \in \{1, \ldots, N-1\}$ uniformly at random such that $\gcd(a, N) = 1$, its order r modulo N is even and N does not divide $a^{r/2} + 1$. Fortunately, this event occurs with a probability of at least $3/8$.

Lemma 8. *Let $N = pq$ be an odd composite number, where p and q are primes. If an integer a is chosen uniformly at random from $\{1, \ldots, N-1\}$ such that $\gcd(a, N) = 1$, then with a probability of at least $3/8$, the order r of a modulo N is even and $a^{r/2} \not\equiv \pm 1 \pmod{N}$.*

A proof of this can be found in [1] (Lemma 9.3).

[1] Umesh Vazirani. Cs 294-2 quantum computation, fall 2004, lecture 9: Shor's factoring algorithm. University of California, Berkeley, Course Lecture Notes, 2004. Accessed: 2026-06-08

4.4 Prerequisites for Shor's order-finding algorithm

Shor's quantum algorithm for order-finding relies on the *Quantum Fourier Transform* (QFT). To understand the QFT, we must first examine some key properties of the *roots of unity*.

4.4.1 Roots of unity

The numbers 1 and -1 satisfy the equation $x^2 - 1 = 0$. Similarly, the numbers $1, i, -1, -i$ satisfy the equation $x^4 - 1 = 0$. One can ask an

analogous question for any n: what are the complex numbers that satisfy the equation $x^n - 1 = 0$? These numbers are called the n-th roots of unity.

Definition 20 (Roots of unity)**.** Let n be a positive integer. A complex number $z \in \mathbb{C}$ is called an n-th root of unity if it satisfies the algebraic equation:

$$x^n - 1 = 0$$

Furthermore, z is called a primitive n-th root of unity if $z^n = 1$ and $z^m \neq 1$ for all integers $1 \leq m < n$.

Note that if ω_n is a primitive n-th root of unity, then $\{1, \omega_n, \omega_n^2, \ldots, \omega_n^{n-1}\}$ is the set of all n-th roots of unity. Geometrically, the n-th roots of unity are equally spaced around the unit circle in the complex plane.

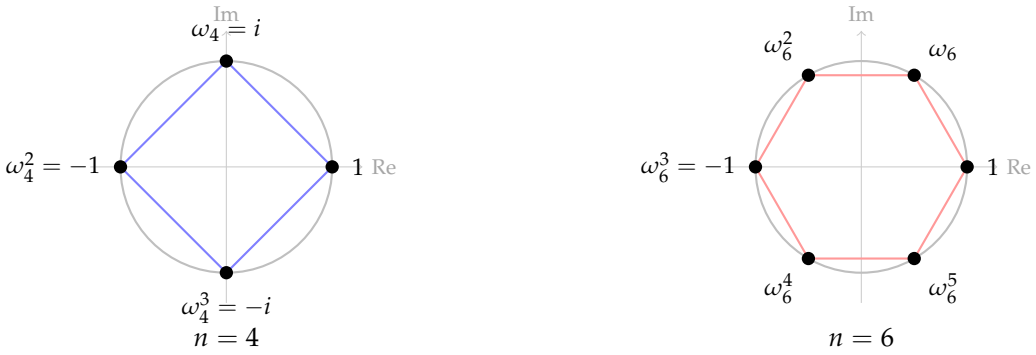

Figure 1. The n-th roots of unity for $n = 4$ (left) and $n = 6$ (right).

Roots of unity satisfy the following useful property.

Theorem 9. *Let $\omega \neq 1$ be an n-th root of unity (not necessarily primitive). Then*

$$\sum_{i=0}^{n-1} \omega^i = 0.$$

Exercise 36 (The QFT matrix is unitary)**.** Let ω be a primitive n-th root of unity. Consider the $n \times n$ matrix Q whose entries are given by:

This matrix defines the Quantum Fourier Transform (QFT).

$$\mathsf{Q}_{i,j} = \frac{1}{\sqrt{n}} \omega^{ij}$$

for $0 \leq i, j \leq n-1$. Show that Q is a unitary matrix (i.e., $\mathsf{Q}^\dagger \mathsf{Q} = \mathsf{I}_n$).

4.4.2 *Continued fractions*

Continued fractions offer an elegant alternative to standard positional decimal notation, representing a real number through a sequence of nested divisions. Formally, a *simple* continued fraction expresses a number x as its integer part plus the reciprocal of another number,

which is itself decomposed into an integer part plus a reciprocal, and so on, yielding an expression of the form

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

where a_0 is an integer and the remaining a_i are positive integers. This can be written compactly as $x = [a_0; a_1, a_2, \ldots]$. For example, the rational number $\frac{45}{16}$ can be progressively decomposed as follows:

$$\frac{45}{16} = 2 + \frac{13}{16} = 2 + \frac{1}{\frac{16}{13}} = 2 + \frac{1}{1 + \frac{3}{13}} = 2 + \cfrac{1}{1 + \cfrac{1}{\frac{13}{3}}} = 2 + \cfrac{1}{1 + \cfrac{1}{4 + \frac{1}{3}}}.$$

The final expression is the continued fraction representation of $\frac{45}{16} = [2; 1, 4, 3]$. This process terminates precisely when the starting value x is rational; for an irrational x it continues indefinitely. Even so, truncating a non-terminating expansion at successive depths yields an increasingly accurate approximation of x. The rational number obtained by keeping only the first $k+1$ partial quotients $a_0, a_1, \ldots, a_k$ is called the *kth convergent*, written $c_k = [a_0; a_1, \ldots, a_k]$. For example, with $x = \frac{45}{16}$ the convergents are

$$c_0 = 2, \qquad c_1 = 3, \qquad c_2 = \tfrac{14}{5}, \qquad c_3 = \tfrac{45}{16},$$

the last of which recovers x itself.

A natural question is the rate of convergence for rational numbers. Suppose $x = p/q$, where both p and q are at most 2^t. Let $x = [a_0; a_1, a_2, \ldots, a_k]$. Can we give a bound on k in terms of t? One can show that $k \leq t$, and the continued fraction expression for x can be computed in time $\mathsf{poly}(t)$. Please refer to [2], Theorem A4.15 and the subsequent discussion for the proof.

[2] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information (10th Anniversary edition)*. Cambridge University Press, 2016

Connection to Shor's algorithm: Recall that the heart of Shor's algorithm is order-finding: computing the order of a number a modulo N, the smallest positive r such that $a^r \equiv 1 \pmod{N}$. The quantum part of the algorithm does not hand us r directly. Instead, it produces an integer y in a register of t qubits with the property that

$$\frac{y}{2^t} \approx \frac{s}{r}$$

for some unknown integer $s \in \{0, 1, \ldots, r-1\}$. To recover r from here, we need some classical post-processing, and this is where continued fractions come in. More formally, Shor's algorithm provides

us with the following guarantee: if (a, N) are the inputs, then for appropriately chosen large (but still polynomial in $\log N$) t, it outputs y (with reasonable probability) such that there exists s satisfying $\gcd(s, r) = 1$ and

$$\left|\frac{y}{2^t} - \frac{s}{r}\right| \leq \frac{1}{2r^2} \tag{4.1}$$

where r is the order of a wrt N. Our goal was to find r. Here, we will use a famous result by Legendre (a proof of this theorem can be found in [3], Theorem A4.16).

[3] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information (10th Anniversary edition)*. Cambridge University Press, 2016

Theorem 10 (Legendre's theorem on continued fraction)**.** *For any $\Phi \in \mathbb{R}$, if*

$$\left|\Phi - \frac{a}{b}\right| \leq \frac{1}{2b^2}$$

for some $a, b \in \mathbb{Z}$ such that $\gcd(a, b) = 1$, then a/b is one of the convergents of the continued fraction expansion of Φ.

Moreover, if $\Phi = y/2^t$, and both a and b are guaranteed to be less than 2^t, then they can be computed in $\mathsf{poly}(t)$ steps. This completes our description of the quantum algorithm for order finding:

1. Given (a, N) such that $\gcd(a, N) = 1$, choose an appropriately large t (polynomial in $\log N$).
2. Using the core quantum component of Shor's algorithm, compute a y such that there exists an $s < r$ satisfying $\gcd(s, r) = 1$ and $\left|y/2^t - s/r\right| \leq 1/(2r^2)$.
3. Using the convergent finding algorithm, find s/r. Output r.

5
Bibliography

[1] Michael Artin. *Algebra*. Birkhäuser, 1998.

[2] Eric Lehman, F. Thomson Leighton, and Albert R. Meyer. *Mathematics for Computer Science*. MIT OpenCourseWare, 2018. Available at `https://courses.csail.mit.edu/6.042/spring18/mcs.pdf`.

[3] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, 2nd edition, 2017.

[4] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information (10th Anniversary edition)*. Cambridge University Press, 2016.

[5] Umesh Vazirani. Cs 294-2 quantum computation, fall 2004, lecture 9: Shor's factoring algorithm. University of California, Berkeley, Course Lecture Notes, 2004. Accessed: 2026-06-08.