Module 1: Foundations

Contents

❖ Foundations 1: Probability and Complex Numbers (1/8)

Humor is the ability to see three sides to one coin - Ned Rorem

1.1 A single coin

You may be familiar with the concept of a **sample space** from probability. This is simply the collection of possible outcomes given a probabilistic procedure. In this course, we will often discuss **state spaces** instead, which are the set of possible states some system can be in. The distinction is that a state represents a dynamic system, where the state is expected to change as time progresses due to some external force applied to it. In our case, this force will be modeled using computational gates, but more on that later.

Suppose we have a biased coin that lands on heads with probability p.

Question 1. What are the possible states that the coin can be in?

An important thing to note is that when we say "state", this corresponds to our best representation of the object, not necessarily the true physical state of the object. Think of it as a mental model for our prediction of the state of the coin. Critically, this means that our mental representation of the coin can change depending on whether we are *looking* at the coin or not.

Question 2. Suppose we took the coin and place it in a box, close it, then give it a good shake. How can we mathematically model and represent the action of shaking the box?

The above examples were instances of some important mathematical tools we will be using.

Definition 1.1 (Probability Vector)**.** A **probability vector** is a vector containing nonnegative real entries that sum to 1. The entries store the probability of seeing the event corresponding to the index.

Definition 1.2 (Stochastic Matrix)**.** A **stochastic matrix** is a matrix with nonnegative elements whose columns add up to 1.

Question 3. Show that multiplying a probability vector by a stochastic matrix always results in another probability vector.

Question 4. Describe the stochastic matrix representing the action of turning the box upside down. We can assume that the action is done stably, meaning that the orientation of the coin will change perfectly with the box.

Question 5. Consider the following coin game using a fair coin (probability of heads is 1/2), where the action will change depending on the state. The action during a single turn is the following:

- If the current state is HEADS: Do a fair coin flip.
- If the current state is TAILS: Turn the coin over.

Describe the stochastic matrix corresponding to this game.

Challenge: If I had someone play this game for me for 100 turns, how would I represent the state of the coin? 1000 turns? Infinite turns? Is there a state the coin will converge to?

Example 1.3. In general, we can describe the entries of a stochastic matrix for a coin by the following:

$$\begin{bmatrix} \mathbb{P}(T_{\text{after}}|T_{\text{before}}) & \mathbb{P}(T_{\text{after}}|H_{\text{before}}) \\ \mathbb{P}(H_{\text{after}}|T_{\text{before}}) & \mathbb{P}(H_{\text{after}}|H_{\text{before}}) \end{bmatrix} \tag{1}$$

This is consistent with our definition, as the columns correspond to conditioned probability distributions implying that the entries are nonnegative and do sum to 1.

1.2 Two coins

How can we extend this mathematical model to a system of two coins? Suppose we have two biased coins, where the state of the first coin is represented by the vector $\begin{bmatrix} a \\ b \end{bmatrix}$ and the state of the second coin is represented by the vector $\begin{bmatrix} c \\ d \end{bmatrix}$. Then,

$$\begin{bmatrix} \mathbb{P}[TT] \\ \mathbb{P}[TH] \\ \mathbb{P}[HT] \\ \mathbb{P}[HH] \end{bmatrix} = \begin{bmatrix} ac \\ ad \\ bc \\ bd \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} \otimes \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a \begin{bmatrix} c \\ d \end{bmatrix} \\ b \begin{bmatrix} c \\ d \end{bmatrix} \end{bmatrix} \tag{2}$$

In the above equation, the symbol "⊗" is called the **tensor product** or Kronecker product and will be the standard way we combine two state spaces.

Question 6. Suppose someone flipped two fair coins (probability of seeing heads is 1/2). How would we represent the full system of the two coins?

Two coins introduces some complexity to our model. It turns out that not all probability vectors with four elements can be described by simply taking the tensor product between two probability vectors with two elements! We will prove this is the case by taking a special example in the following question.

Question 7. Consider the following probability vector:

$$B = \begin{bmatrix} 1/2 \\ 0 \\ 0 \\ 1/2 \end{bmatrix} \tag{3}$$

Prove that this vector cannot be constructed by taking the tensor product between two probability vectors representing coins.

Question 8. Suppose both of the coins begin in the state tails. Is there a sequence of actions on the individual coins (think stochastic matrices) such that the final state of the two coins will be B defined in equation (3)?

We will call states that can't be decomposed into a tensor product of smaller states a **correlated state**. Why correlated? If I told you what the state of the first coin, you will immediately know what the state of the second coin is!

1.3 Inference

If someone gave us a coin and asked us to determine if it was weighted, how would we do so? Could we tell them what the exact weight is? This is a core problem in quantum computing that gets quite challenging, so it'll help to start thinking about this early on. Here we will consider the simple case of the coin, a.k.a. a distribution with two possible outcomes.

1.4 Some Review

Let's review some key ingredients we will need for our proofs. Once we are done with a sequence of steps, we can think of our state as representing a sample space. We would like to sample from this state to determine what the bias of the coin p is.

Definition 1.4. A **random variable** is a function $X : S \to \mathbb{R}$ where S is our sample space. That is, with every $x \in S$ there is an associated real number $X(x)$.

The **expected value** of a random variable X is defined to be

$$\mathbb{E}(X) = \sum_{x \in S} p(x)X(x) \tag{4}$$

where $p(x)$ is the probability that x occurs.

Question 9. Define the random variable X over the sample space of a coin as follows:

$$X(T) = 0 \tag{5}$$

$$X(H) = 1 \tag{6}$$

Suppose our coin has a probability of 0.8 to be heads. What is $\mathbb{E}(X)$?

Theorem 1.5 (Linearity of Expectation)**.** If we have two random variables X and Y, then

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) \tag{7}$$

and

$$\mathbb{E}(cX) = c\mathbb{E}(X) \tag{8}$$

for any constant $c \in \mathbb{R}$.

Question 10. Prove the linearity of expectation.

Question 11. Is it true that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ for two random variables X and Y?

Definition 1.6. The **variance** of a random variable X is defined as

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}[f^2] - \mathbb{E}[f]^2. \tag{9}$$

The variance describes how closely the value of the random variable clusters around its expected value.

The **standard deviation** $\sigma(X)$ of a random variable is defined to be the square root of $\text{Var}(X)$.

Question 12. Show that the two definitions of variance stated above are equivalent.

Given some random variable, we would like to bound the probability that it deviates from some value such as its mean. Why does this help us with inference? Suppose we would like to estimate the probability that a coin is heads. If we can show that the probability that over 300 coin tosses, the probability that our sample mean is significantly different from the true mean, we can confidently use our sample mean as the estimate of the true mean. The key tool computer scientists use for this task is called a **Chernoff bound**. Here we will go over the proof of weaker tools for a similar task, and simply state the Chernoff bound.

Theorem 1.7 (Markov's Inequality)**.** Let $X : S \to \mathbb{R}$ be a nonnegative random variable. Then, for any $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}. \tag{10}$$

Proof.

Markov's inequality is quite loose, and doesn't tell us how many samples we would need for a good approximation. Consider the following problem:

Question 13. Suppose we have a weighted coin that has probability 0.2 to land on heads. If we tossed the coin 20 times, find a bound for the probability that 16 of them were heads using Markov's inequality.

It turns out that you can't do much better than Markov's inequality in general if you only know the expected value. However, if we use the variance, we can get a tighter bound.

Theorem 1.8 (Chebyshev's Inequality)**.** Let $X : S \to \mathbb{R}$ be a random variable with expectation $\mathbb{E}(X)$ and variance $\text{Var}(X)$. Then, for any $a \in \mathbb{R}$:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}. \tag{11}$$

Proof.

Question 14. Consider the same scenario as question 13. What does Chebyshev's inequality say about the probability of this event? The variance of a Binomial distribution is given by $np(1-p)$.

Theorem 1.9 (Chernoff-Hoeffding Bound)**.** Let $X_1, \ldots, X_n$ be independent random variables such that $a_i \leq X_i \leq b_i$. Consider the sum of these random variables,

$$S_n = X_1 + \cdots + X_n. \tag{12}$$

Then, for all $\epsilon > 0$,

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq \epsilon) \leq 2\exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \tag{13}$$

1.5 Complex Numbers

The shortest path between two truths in the real domain passes through the complex domain.

- Jacques Hadamard

Before diving into our exploration of quantum states, it'll be helpful to review some important definitions and properties about complex numbers, as well as the vectors and matrices that have complex entries.

Definition 1.10 (Complex Number (Standard Representation))**.** A **complex number** α is a number that can be written as

$$\alpha = a + bi \tag{14}$$

for two real numbers a and b, and i is defined to be the constant satisfying $i^2 = -1$. This is the **standard representation** (or **standard form**) of expressing a complex number. The set of all complex numbers will be written as $\mathbb{C}$.

Every complex number has a **complex conjugate**. The complex conjugate of $\alpha = a + bi$ is

$$\alpha^* = a - bi. \tag{15}$$

Definition 1.11 (Norm of Complex Number)**.** The **norm** $|\alpha|$ of a complex number α is defined as

$$|\alpha| = \sqrt{a^2 + b^2}. \tag{16}$$

This is the "size" of the complex number, and we can see why this is true by looking at the complex plane. Because of this, $|\alpha| > 0$ for any complex number, and the only time $|\alpha| = 0$ is when $\alpha = 0$.

We will also often be using the phase representation of a complex number.

Definition 1.12 (Complex Number (Phase Representation))**.** On the complex plane, we can also represent the number using the counterclockwise angle θ from $1 + 0i$, and its norm $|\alpha|$. That is,

$$\alpha = |\alpha|(\cos\theta + i\sin\theta) = |\alpha|e^{i\theta} \tag{17}$$

where the last equality uses the identity $e^{i\theta} = \cos\theta + i\sin\theta$.

Question 15. What is the complex conjugate of $\alpha = |\alpha|e^{i\theta}$ in the phase representation?

Question 16. Compute the norm of $\alpha = 3 + 7i$, $\beta = -2 - 3i$, $\gamma = \frac{1}{\sqrt{3}} - \sqrt{\frac{2}{3}}i$.

Question 17. Show that $|\alpha| = \sqrt{\alpha^* \alpha} = \sqrt{\alpha \alpha^*}$. Try computing the norm of α from the above example using this method.

Question 18. Express the complex number with phase representation $7 \cdot e^{i5\pi/6}$ in the standard representation.

1.6 Summary and future directions

We defined a probability vector to be a vector that models a distribution over states. We did this by assigning two quite natural constraints,

1. the entries must sum to 1, and
2. the entries must be nonnegative real numbers.

As mathematicians, we like to generalize requirements on interesting objects. How would we generalize the above definition?

The first constraint is equivalent to saying that the $L1$-norm of the vector must be 1 (we will review norms, but if this doesn't ring a bell you should do a quick google search on the definition). This, however, is not the standard norm we study in linear algebra! We are more familiar with measuring the length of a vector by the $L2$-norm. Maybe we can require the vectors to be unit vectors in the standard $L2$-norm.

Once we begin considering $L2$-norms, the entries don't need to be nonnegative real numbers, since we will be squaring them anyways. Maybe we can use negative numbers, or even complex numbers!

It turns out that these two generalizations for a probability vector are exactly what physicists use to describe quantum states, and more importantly what we will be using to represent the states of quantum computers. We will show that this is a good generalization as many of the questions we answered here will work for quantum states as well.

❖ Foundations 2a: Introduction to Qubits (1/12)

2.1 A Single Qubit

We are now ready to begin our discussion of quantum states. We started our discussion around probability vectors representing two events, which will be a good starting place for discussing quantum states. Let's start by defining a quantum bit or qubit using the same language.

Definition 2.1. A **qubit** is an object which can be represented using a *unit vector* with complex **amplitudes** α_0 and α_1 as

$$|\psi\rangle = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}, \tag{18}$$

where we say that α_i is the amplitude corresponding to the event i for $i \in \{0, 1\}$. The notation $|\psi\rangle$ is read as "ket" "psi".

Two of the most important states that we will be using throughout the course are the **standard basis states**, defined as

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{19}$$

These are the vectors corresponding to the two primary states that the system can be in. We can use linearity of vectors to write equation (18) as a linear combination of the standard basis states:

$$|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle . \tag{20}$$

When both coefficients are nonzero, we say that the state $|\psi\rangle$ is in **superposition**.

We are now equipped with the language to represent the state of a qubit. How do we interpret this state? A crucial operation in quantum computing is measurement, which, for our purposes is the way we read out the result of a quantum algorithm. If we measure $|\psi\rangle$ in the standard basis,

- with probability $|\alpha_0|^2$: we *observe* the outcome $|0\rangle$, and the qubit *collapses* to $|0\rangle$.
- with probability $|\alpha_1|^2$: we *observe* the outcome $|1\rangle$, and the qubit *collapses* to $|1\rangle$.

Similar to the case of the probability vector, *observation* collapses the state to the one that we observe. The key difference that makes quantum states special is the fact that there are physical particles which can truly represent superposition, whereas our discussion around probability vectors was slightly superficial.

Question 19. Suppose we have the state

$$|\phi\rangle = \left(\frac{1}{\sqrt{6}} - i\frac{1}{\sqrt{6}}\right)|0\rangle + \left(\frac{1}{\sqrt{3}} + i\frac{1}{\sqrt{3}}\right)|1\rangle . \tag{21}$$

What is the probability of measuring $|0\rangle$, and what is the state after the measurement?

2.2 Multiple Qubits Sneak Peek

To represent a probability distribution over more than two states, we simply use a probability vector with the number of states we need.

Question 20. Suppose we had n bits available. How many states can we represent?

The probability vector representing the distribution over all n-bit strings would look something like this, where p_x is the probability that event x occurs:

$$\text{state} = \begin{bmatrix} p_0 \\ p_1 \\ \vdots \\ p_{2^n-1} \end{bmatrix}. \tag{22}$$

By definition, we require that the entries are nonnegative and sum up to 1, so that they represent a proper probability distribution.

We can generalize this the same way we did for a single qubit. Instead of probabilities for events occurring, each event x has an associated complex number α_x called its **amplitude**. As a vector, this would look like

$$|\psi\rangle = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{2^n-1} \end{bmatrix}. \tag{23}$$

Equivalently, we will often use ket notation to express the state, which for the above vector would look like:

$$|\psi\rangle = \alpha_0 |00\ldots00\rangle + \alpha_1 |00\ldots01\rangle + \cdots + \alpha_{2^n-1} |11\ldots11\rangle . \tag{24}$$

Here, we used the shorthand

$$|00\rangle = |0\rangle \otimes |0\rangle . \tag{25}$$

Again, we require that the vector is a unit vector: $\sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1$. This ensures that the squared norm of the amplitudes form a probability distribution.

Question 21. Write down a 3-qubit state where the probability of measuring each qubit is equal. Try to express your answer both as a statevector and in ket notation.

2.3 Transformations

Quantum algorithms have three main components:

1. Store quantum information (statevector)
2. Manipulate quantum information (unitary transformations)
3. Extract some output (quantum measurement)

We've seen examples of what 1 and 3 look like from a theoretical point of view, so here we will briefly discuss 2. The manipulation of quantum information can be thought of as a transformation from one quantum state to a new quantum state, which can be expressed in vector form as

$$\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} \rightarrow \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}. \tag{26}$$

One requirement we have for these transformations is that they be linear. This means that if we know what a transformation does for all basis vectors, we will know how *any* vector will be transformed. We will review this more carefully in the next section, so here we explore an example for single qubit states.

Question 22. Suppose we have a linear transformation T that acts as follows on the standard basis states:

- $T|0\rangle = T\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix}$
- $T|1\rangle = T\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$

What is the action of T on the state $|\psi\rangle := \gamma_0|0\rangle + \gamma_1|1\rangle$?

We can always represent linear transformation with matrices. In our case, we further have the requirement that the transformation should map unit vectors to unit vectors. This property is satisfied by the family of **unitary matrices**.

Definition 2.2 (Unitary Transformation)**.** Any linear transformation that preserves $L2$-norm is **unitary** (think takes unit vectors to unit vectors), and a matrix that represents such a transformation is a **unitary matrix**.

❖ Foundations 3a: Linear Algebra (1/12)

Hilbert space is a big place.

- Carlton Caves

3.1 Vector Spaces

In linear algebra, we are interested in studying vector spaces.

Definition 3.1 (Vector Space)**.** A **vector space** is a set of elements that is closed under linear combinations. A **linear combination** is a combination of vectors via vector addition and scalar multiplication.

The primary focus of this course will be the **complex vector space** of N dimensions, which will be referred to via the short hand $\mathbb{C}^N$. You may also see me (and others) refer to the "Hilbert space" of quantum states. These are the same thing, as a Hilbert space can be thought of as a vector space where you can take inner products. Elements of $\mathbb{C}^N$ are vectors of the form

$$|v\rangle = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}. \tag{27}$$

You may be familiar with using $\vec{v}$ to represent vectors, but here we will use $|v\rangle$ to represent column vectors.

Question 23. Verify that $\mathbb{C}^N$ is indeed a vector space. I.e., are the elements closed under linear combinations.

3.2 Span and Linear Independence

Definition 3.2 (Span)**.** The **span** of a set of N vectors $\{|\psi_1\rangle, \ldots, |\psi_N\rangle\}$ is the set of all linear combinations of $|\psi_1\rangle, \ldots, |\psi_N\rangle$, i.e. the set of all states that can be written as

$$c_1 |\psi_1\rangle + \cdots + c_N |\psi_N\rangle \tag{28}$$

for all complex scalars $c_1, \ldots, c_N \in \mathbb{C}$.

Question 24. If the following statement is true, prove it. If not, provide a counterexample:

For any pair of vectors $|v\rangle, |w\rangle$ in $\mathbb{R}^2$, the span of $|v\rangle$ and $|w\rangle$ is all of $\mathbb{R}^2$. In other words, any two pair of vectors spans the entire space.

Definition 3.3 (Linearly Independent Set of Vectors)**.** Let $B = \{|\psi_1\rangle, \ldots, |\psi_N\rangle\}$ be a set of vectors in $\mathbb{C}^N$. We say that this set of vectors is **linearly independent** if

$$c_1 |\psi_1\rangle + \cdots + c_N |\psi_N\rangle = 0 \tag{29}$$

if and only if $c_i = 0$ for all i.

The above definition is equivalent to saying that no basis vector can be written as a linear combination of the other basis vectors. Equivalently, we say that a set of vectors is independent if for any $|\phi\rangle \in \mathbb{C}^N$, there is a *unique* set of scalars $c_1, \ldots, c_N \in \mathbb{C}^N$ such that

$$c_1 |\psi_1\rangle + \cdots + c_N |\psi_N\rangle = |\phi\rangle . \tag{30}$$

3.3 Inner Products and Bases

We can equip a vector space with an inner product, which is an operation that maps two vectors to a scalar value. We will refer to a vector space with an inner product a **Hilbert space**.

Definition 3.4 (Inner Product ($\mathbb{C}^N$))**.** Let $|\psi\rangle = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}$ and $|\phi\rangle = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}$ be elements of $\mathbb{C}^N$. Then the **inner product** between $|\psi\rangle$ and $|\phi\rangle$ is $\sum_i \alpha_i^* \beta_i$. In matrix product form, an equivalent way to write this is

$$\begin{bmatrix} \alpha_1^* & \cdots & \alpha_N^* \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} = \sum_i \alpha_i^* \beta_i. \tag{31}$$

In ket notation, we write the **dual** of a complex vector $|\psi\rangle$ as $\langle\psi| := \begin{bmatrix} \alpha_1^* & \cdots & \alpha_N^* \end{bmatrix}$, read as **bra** psi. Using this notation, the inner product is often written as $\langle\psi|\phi\rangle$. We refer to the notation of writing vectors with these angle brackets as **bra-ket notation**.

Question 25. Show that $\langle\psi|\phi\rangle = \langle\phi|\psi\rangle^*$.

Question 26. What happens when we take the inner product of a vector with itself? Does it relate to a quantity about vectors you've seen before?

Definition 3.5 (L2-norm)**.** For a vector $|\psi\rangle \in \mathbb{C}^N$, the **L2-norm** of $|\psi\rangle$ denoted is

$$|| \, |\psi\rangle \, || := \sqrt{\langle\psi|\psi\rangle}. \tag{32}$$

In this course, when we say norm, we will be referring to the L2-norm of the vector unless otherwise stated. If the norm of a vector is 1, we say that vector is a **unit vector**.

Question 27. What is the norm of $|\psi\rangle = (2+i)|0\rangle + (3-2i)|1\rangle$?

Definition 3.6 (Orthogonality)**.** Given two vectors $|\psi\rangle$ and $|\phi\rangle$ in $\mathbb{C}^N$, we say that they are **orthogonal** if $\langle\psi|\phi\rangle = 0$.

The inner product is a useful metric in defining a notion of similarity between two vectors. A high inner product between two vectors is colloquially said to have high overlap, and they point in similar directions.

Definition 3.7 (Orthogonal Basis)**.** If a set of N vectors $B = |\psi_1\rangle, \ldots, |\psi_N\rangle$ in $\mathbb{C}^N$ is mutually orthogonal (i.e., if $i \neq j$ then $\langle\psi_i|\psi_j\rangle = 0$), we say that B forms an **orthogonal basis** for $\mathbb{C}^N$.

Furthermore, if every vector $|\psi_i\rangle$ is also a unit vector, we call it an **orthonormal basis**.

Question 28. Write a set of orthonormal basis vectors for $\mathbb{R}^2$ besides the standard basis and draw it on the plane. Find an orthonormal basis for $\mathbb{C}^4$ where one of the vectors is $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$

3.4 Summary

We have now covered the main foundational mathematical concepts we will be using to build our understanding of quantum computing. One thing I have really enjoyed about quantum computing is that it gave me a new way to visualize and understand the above tools, which you may have felt were quite abstract in your preliminary courses. I hope this new angle will give you a new appreciation and understanding of these tools. Next week we will start looking at small quantum systems and get familiar with the circuits we will use to prove ideas about the limits of information and construct algorithms.