CS 269Q: Section 1

Multilinear Algebra and Quantum Mechanics

Jon Braatz
jfbraatz@stanford.edu

Duligur Ibeling
duligur@stanford.edu

April 5, 2019

Table of Contents

Plan

1. Notational Bookkeeping for Probability Distributions
   - Key Idea: Use (multi)linear algebra to keep track of probabilities
2. Quantum States/Amplitude Distributions/Wavefunctions
   - Key Idea: Using amplitudes instead of probabilities and allowing measurements in different bases turns Probability Theory into Quantum Mechanics

Probability Distributions

- A probabilistic system may be in an uncertain "probabilistic state" such that after we "measure" it we find that it is in exactly one of n mutually exclusive "definite states".
- Each definite state is associated with a real number $p_i \in [0, 1]$ such that the probability a measurement finds the system to be in state i is p_i.
- Since the definite states are mutually exclusive and we're considering all possible definite states, we have $\sum_{i=1}^{n} p_i = 1$.
- Measurement turns any probabilistic state into a definite state (measuring multiple times in a row gives the same answer every time)
- We want to develop a convenient notation for manipulating these (classical) probability distributions.

Concrete Example: Loaded Dice

- For example, our probabilistic system might be a loaded die with faces labelled 1 through 6 and some probability p_i for each face.
- The definite states are the faces, and we can put the die into a probabilistic state by rolling it without looking at the result.
- We measure it by looking at the number on the top face, after which the probabilistic state "collapses" into a definite state
- It's definite because we can look at the number on the top face as many times as we like and we'll observe the same number every time (so long as nothing else perturbs it somehow).
- What kinds of questions can we ask about such experiments?

Questions We Can Ask About Loaded Dice

- We note that each definite state is naturally associated with a real number, namely the number on the face. We call these numbers "observables" of the system.
- What are the expected values of these observables? What are the expected values of any function of these observables? What about standard deviations?
- What about probabilities and expected values of observables corresponding to outcomes of multiple rolls?
- What about for rolls of multiple differently loaded dice?
- How do those answers change when we shift the distributions around?
- We know how to solve simple probability word problems like these, but we'd like to develop a notation to do all the bookkeeping for us.

Creating Our Bookeeping Notation

- We've been talking about particular probabilistic systems and states they can be in, so our notation should allow us to easily name these entities.
- If our system is named ψ, we'll denote by $|\psi\rangle$ the probabilistic state that it's in.
- Definite states are particular kinds of probabilistic states where the probabilities of being measured in other states is 0, so to be consistent we should use the same notation for them.
- Let's denote the name of the i^{th} definite state simply i, so that the state itself is denoted $|i\rangle$.
- The state $|\psi\rangle$ is completely specified by the probabilities p_i, and we want to keep track of which definite state each probability corresponds to, so we'll tag each p_i with $|i\rangle$ and represent $|\psi\rangle$ by a formal sum of these tagged probabilities:

$$|\psi\rangle = p_1|1\rangle \oplus p_2|2\rangle \oplus \ldots \oplus p_n|n\rangle$$

More Bookkeeping

$$|\psi\rangle = p_1|1\rangle \oplus p_2|2\rangle \oplus ... \oplus p_n|n\rangle$$

- At this point, this is a purely syntactic expression. We're using juxtaposition of probabilities and definite states as a way to tag those probabilities with their corresponding definite states, and we're just using $\oplus$ to denote an unordered sequence of these tagged values.
- Think of it as notation for a key-value store mapping definite states to probabilities
- We will end up adding useful structure to this notation by introducing simplification/rewrite rules that "implement" the common manipulations of probability distributions.
- Now that we have a minimum viable representation of the state of system ψ, we can start asking question about it and seeing if we can encode the answers with additional notation and rewrite rules.

Asking and Answering Questions With Notation

$$|\psi\rangle = p_1|1\rangle \oplus p_2|2\rangle \oplus \ldots \oplus p_n|n\rangle$$

- The simplest thing we can do with a distribution is to get out one of the probabilities of our choosing. So how should we denote "doing a lookup" on this key value store?
- Syntactically we can represent the question "What is the probability of outcome j?" by $\langle j|$, and ask a state this question by juxtaposition on the left, so that we can say that $\langle j|\psi\rangle = p_j$.
- Let's consider how that would work for definite states: we know that the answer p_j for state $|i\rangle$ is 1 if $i = j$ and 0 otherwise, by the definition of definite states.
- So we have our first rewrite rule:

$$\langle j|i\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Asking and Answering Questions With Notation

- We want to be able to ask question $\langle j|$ of any general probabilistic state $|\psi\rangle$, but our first rewrite rule only applies to asking such questions of definite states.
- We want to add a rewrite rule that allows us to say:

$$\langle j|\psi\rangle = \langle j|(p_1|1\rangle \oplus \ldots \oplus p_j|j\rangle \oplus \ldots \oplus p_n|n\rangle) = p_j$$

- We want every term besides the one with $|j\rangle$ to disappear.
- We sense that our first rewrite rule might help us construct a new rewrite rule, since it implements a kind of selection operation.

Asking and Answering Questions With Notation

- If we apply the first rewrite rule with $\langle j|$ everywhere we can to $|\psi\rangle$, we get:

$$p_1\langle j|1\rangle \oplus \ldots \oplus p_j\langle j|j\rangle \oplus \ldots \oplus p_n\langle j|n\rangle = p_1 0 \oplus \ldots \oplus p_j 1 \oplus \ldots \oplus p_n 0$$

- Defining the juxtaposition of the numbers in each term to be scalar multiplication and the $\oplus$ operation to be the usual addition operator $+$ would allow us to simplify this expression to be p_j, so we'll make those definitions and consider juxtaposition and $\oplus$ to be multiplication and $+$ from now on.
- We'll make the final connection allowing us to equate $\langle j|\psi\rangle$ with p_j by defining $\langle j|\psi\rangle$ to be the expression on the left hand side of the above equation so that we have:

$$\begin{aligned}\langle j|\psi\rangle &= \langle j|(p_1|1\rangle + p_2|2\rangle + \ldots + p_n|n\rangle) \\ &= p_1\langle j|1\rangle + \ldots + p_j\langle j|j\rangle + \ldots + p_n\langle j|n\rangle \\ &= p_1 0 + \ldots + p_j 1 + \ldots + p_n 0 = p_j\end{aligned}$$

Linear Algebra = Notational Bookkeeping

- I hope it's clear by now that the rewrite rules we came up with are the notation and axioms of a finite dimensional vector space with standard Euclidean metric. So from now on, regular vector space rules apply.
- The definite state tags we started with can be interpreted as basis column vectors, and the $\langle j|$'s as basis row vectors.
- A probabilistic state is a linear combination of the basis outcome vectors with coefficients being the corresponding probabilities.
- The probability of a definite state is obtained by forming the inner product of the probabilistic state vector with that definite state basis vector: $p_j = \langle j|\psi\rangle$.
- If we denote by $\langle u|$ the sum of the basis row vectors $\sum_{i=1}^{n} \langle i|$, then the normalization condition that probabilities sum to 1 is written as $\langle u|\psi\rangle = 1$.
- This is called Dirac notation, and we call column vectors "kets" or "ket vectors" and row vectors "bras" or "bra vectors".

Dirac Notation for Linear Operators

- We can use outer products as an alternative notation to matrices for representing linear transformations.
- The matrix $n \times n$ matrix A with element a_{ij} at row i column j can be represented using Dirac notation as $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} |i\rangle\langle j\rangle$.
- Then if we have a state $|\psi\rangle = \sum_{i=1}^{n} \psi_i |i\rangle$, we have:

$$
\begin{aligned}
A|\psi\rangle &= (\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} |i\rangle\langle j|)(\sum_{k=1}^{n} \psi_i |k\rangle) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ij} \psi_k |i\rangle\langle j|k\rangle = \sum_{i=1}^{n} (\sum_{k=1}^{n} a_{ik} \psi_k) |i\rangle
\end{aligned}
$$

- The coefficients of the result are what you would get with matrix-vector multiplication.

Observables and Expectation Values

- We can change $|\psi\rangle$ to a different but still well-normalized distribution by acting with any linear operator S such that $\langle u|S|\psi\rangle = \langle u|\psi\rangle$. This will be true if $\langle u|S = \langle u|$, which means that S is a stochastic matrix (columns sum to 1).
- If state $|i\rangle$ is associated with an real observable m_i, then its expectation value is defined to be $\sum_{i=1}^{n} p_i m_i$.
- This is similar to the sum for the normalization condition, but the coefficients of each basis vector $|i\rangle$ has been multiplied by m_i.
- Algebraically, multiplying each component of the state vector by m_i is a linear operator that we can call M, so $M|i\rangle = m_i|i\rangle$ and the expectation value of the observable is $\langle M\rangle = \langle u|M|\psi\rangle$.

Observables as Self-Adjoint Operators

- M's defining equation $M|i\rangle = m_i|i\rangle$ is an eigenvalue equation. M has a set of orthonormal eigenvectors $|i\rangle$, i.e. the definite states, and real eigenvalues m_i.
- Since definite states are eigenvectors of the observable M, what we've been calling "definite states" are usually referred to as "eigenstates of M", but they are semantically identical concepts.
- Observables with classical probability distributions are associated with self-adjoint linear operators, but those operators are diagonal in the basis of definite states, which is the only basis we ever care about classically.

Adjoints of Operators

- Reminder: In usual vector notation, the adjoint of an $n \times n$ matrix A is the matrix $A^\dagger$ such that for any vectors $x, y \in \mathbb{R}^n$ we have $(Ax) \cdot y = x \cdot (A^\dagger y)$. Using transposes instead of dot product notation, this says $(Ax)^\top y = x^\top A^\top y = x^\top A^\dagger y$. So $A^\dagger = A^\top$.
- In general, the class of linear operators with real eigenvalues and orthonormal eigenvectors is the class of self-adjoint operators with respect to the inner product being used.
  - For real L2 norm these are the symmetric operators
  - For complex L2 norm these are the Hermitian operators
- The fact that an operator has real eigenvalues and orthonormal eigenvectors if and only if that operator is self-adjoint is (a special case of) the Spectral Theorem.

Composite Systems

- We're well equipped to manipulate probability distributions of systems like biased coins, but would we handle multiple coins?
- Let's consider two coins $|\psi\rangle = \psi_T|T\rangle + \psi_H|H\rangle$ and $|\phi\rangle = \phi_T|T\rangle + \phi_H|H\rangle$.
- The definite states for each coin are $|H\rangle$ and $|T\rangle$, and we'll denote the four definite states of the composite system as $|T\rangle|T\rangle$, $|T\rangle|H\rangle$, $|H\rangle|T\rangle$, and $|H\rangle|H\rangle$, with the ψ outcome on the left and ϕ on the right.
- For two independent sytems we use the rule that probabilities multiply to write the composite state as $\psi_T\phi_T|T\rangle|T\rangle + \psi_T\phi_H|T\rangle|H\rangle + \psi_H\phi_T|H\rangle|T\rangle + \psi_H\phi_H|H\rangle|H\rangle$.

Tensor Products for Joint States

$$\psi_T\phi_T|T\rangle|T\rangle + \psi_T\phi_H|T\rangle|H\rangle + \psi_H\phi_T|H\rangle|T\rangle + \psi_H\phi_H|H\rangle|H\rangle$$

- The above expression is syntactically what we want the result to be when we combine states $|\psi\rangle$ and $|\phi\rangle$.
- "Multiplying" $|\psi\rangle$ and $|\phi\rangle$ and simplifying the way you would expect is the natural definition to make that yields the expression we're after:

$$\begin{aligned}|\psi\rangle|\phi\rangle &= (\psi_T|T\rangle + \psi_H|H\rangle)(\phi_T|T\rangle + \phi_H|H\rangle)\\ &= \psi_T\phi_T|T\rangle|T\rangle + \psi_T\phi_H|T\rangle|H\rangle + \psi_H\phi_T|H\rangle|T\rangle + \psi_H\phi_H|H\rangle|H\rangle\end{aligned}$$

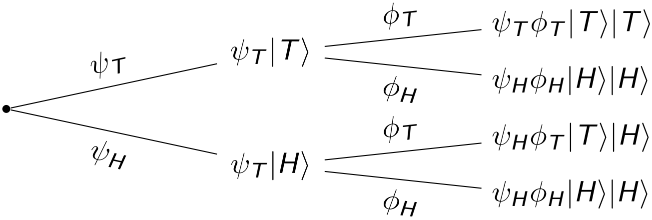
- This product operation between ket vectors is called the tensor product (or sometimes the Kronecker product), and it's sometimes denoted by $\otimes$, but we will usually denote the tensor product of, for example, $|0\rangle$ and $|1\rangle$ by either $|0\rangle|1\rangle$ or $|01\rangle$. The tensor product of two bra vectors is analogous.
- It's a way to keep track of element-wise multiplications.

Tensor Products With Vector Notation

- Using column vector notation for $|\psi\rangle$ and $|\phi\rangle$, we can compute their tensor product as follows:

$$\begin{pmatrix}\psi_T\\ \psi_H\end{pmatrix} \otimes \begin{pmatrix}\phi_T\\ \phi_H\end{pmatrix} = \begin{pmatrix}\psi_T\begin{pmatrix}\phi_T\\ \phi_H\end{pmatrix}\\ \phi_H\begin{pmatrix}\phi_T\\ \phi_H\end{pmatrix}\end{pmatrix} = \begin{pmatrix}\psi_T\phi_T\\ \psi_T\phi_H\\ \psi_H\phi_T\\ \psi_H\phi_H\end{pmatrix}$$

- This notation exactly mirrors how we build probability trees:



Tensor Products With Matrix Notation

- The tensor product $\otimes$ of two linear operators A and B is defined to be the bilinear operation that makes the following identity hold:

$$(A|\psi\rangle) \otimes (B|\phi\rangle) = (A \otimes B)(|\psi\rangle|\phi\rangle)$$

- It works analogously to vector tensor products, as the following example illustrates:

$$\begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \otimes \begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix} = \begin{pmatrix} 0\begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix} & 1\begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix} \\ 2\begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix} & 3\begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} 0\cdot 4 & 0\cdot 5 & 1\cdot 4 & 1\cdot 5 \\ 0\cdot 6 & 0\cdot 7 & 1\cdot 6 & 1\cdot 7 \\ 2\cdot 4 & 2\cdot 5 & 3\cdot 4 & 3\cdot 5 \\ 2\cdot 6 & 2\cdot 7 & 3\cdot 6 & 3\cdot 7 \end{pmatrix}$$

Plan

1. Notational Bookkeeping for Probability Distributions
   - Key Idea: Use (multi)linear algebra to keep track of probabilities
2. Quantum States/Amplitude Distributions/Wavefunctions
   - Key Idea: Using amplitudes instead of probabilities and allowing measurements in different bases turns Probability Theory into Quantum Mechanics

What's Different in Quantum Mechanics?

- All of the math and notation we developed for manipulating probability distributions using multilinear algebra carries over to the quantum setting.
- Instead of real probabilities between 0 and 1 that sum to 1, we use complex numbers called amplitudes whose norms-squared are interpreted as probabilities that sum to 1.
- The normalization condition for an n-state quantum state $|\psi\rangle = \sum_{i=1}^{n} \psi_i |i\rangle$ is $\sum_{i=1}^{n} \psi_i^* \psi_i = 1$
- Classically we had $\langle u|\psi\rangle = 1$, but to add the extra ψ_i^* factors we need to scale the i^{th} coefficient of $\langle u|$ by ψ_i^*.
- Since this new bra vector depends on the coefficients of $|\psi\rangle$, we'll call it $|\psi\rangle^\dagger = \langle\psi| = \sum_{i=1}^{n} \psi_i^* \langle i|$ so that the normalization condition is $\langle\psi|\psi\rangle = 1$. This allows us to define a complex L2 inner product of states $|a\rangle$ and $|b\rangle$ as $(|b\rangle)^\dagger |a\rangle = \langle b|a\rangle = \sum_i a_i b_i^*$.

QM Continued

- Whereas classically the normalization-preserving operators were stochastic matrices, in QM they are the matrices U satisfying $\langle\psi|U^\dagger U|\psi\rangle = \langle\psi|\psi\rangle$, i.e those operators such that $U^\dagger U = (U^*)^\top U = I$, where I is the identity operator.
- These matrices are called unitary matrices, and are the complex analogue of orthogonal matrices.
- Operators corresponding to observables are now Hermitian ($H^\dagger = (H^*)^\top = H$) rather than symmetric, and expectation values of an observable M is now $\langle\psi|M|\psi\rangle$ rather than $\langle u|M|\psi\rangle$
- As in the classical case, when a measurement is performed, the state collapses to eigenstate $|i\rangle$ of the matrix with some probability, which is calculated in the quantum case as $|\langle i|\psi\rangle|^2$. Continuing to measure the observable always yields the same outcome.

Different Eigenbases and Superposition

- The weirdest thing about quantum mechanics that breaks drastically from the classical case is the existence of observables with different sets of eigenstates.
- Algebraically, this means that whereas classically the observable matrices were all diagonal and all commuted with each other ($[A, B] = AB - BA = 0$), quantum observable operators need not commute with one another.
- Measuring the state in one way might put the state in one state out of a certain set of definite outcome states corresponding to that type of measurement. The set of outcome states corresponding to different kinds of measurements might be completely different.
- An eigenstate of one observable might be a superposition of eigenstates of another.
- Measure property X and get result x_1, measure another property Z and get result z. Measuring X again might yield a result $x_2 \neq x_1$. This is very weird.

CS 269Q Section 1: (Multi)linear Algebra and Quantum Mechanics

Jon Braatz, Duligur Ibeling

Two physical principles

- Principle of relativity (Einstein)
  - All inertial observers equivalent (Galileo)
  - No information transmitted superluminally
- Principle of complementarity (Bohr)
  - Complementary properties not all known together
  - I.e., quantum mechanics (matrices don’t commute)

Now we’ll see a situation in which the 2nd seems to contradict the 1st.

EPR pairs (Bell states)

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

- Suppose 2 qubits are prepared as above, “entangled”
- One sent to Romeo and the other to Juliet
  - Zillion miles apart
- Question: what happens to Juliet’s bit if Romeo measures his to be 0? 1?

EPR pairs (Bell states)

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

- Answer: by QM, post measurement, Juliet's bit is instantly *in the same state* as Romeo's
- Telegraphy. But they're so far apart. Superluminal flirtation. Contradiction.
- Why not contradiction?
  - Their love is so strong that nothing can keep them apart
  - JK. **This isn't legitimately transferring info.** Romeo has no control over the measurement, both outcomes happen with 50% chance. Can't use to send a love letter.
  - But we will presently see how entanglement can be exploited

Superdense coding

- We model each message in their courtship as 1 of 4 things:
  - Interest
  - Disinterest
  - Fluff (chatting about their day)
  - An invitation
- Thus classically, how many bits for each message?
  - 2 bits because 2^2 = 4
  - E.g. 00, 01, 10, 11 respectively
- We'll show that by exploiting entanglement, we can get it done with 1 qbit per message

Superdense coding

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

$$|\Phi^-\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$$

$$|\Psi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}$$

$$|\Psi^-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

- Question: compute the following.

$$I_1 |\Phi^+\rangle =$$

$$X_1 |\Phi^+\rangle =$$

$$iY_1 |\Phi^+\rangle =$$

$$Z_1 |\Phi^+\rangle =$$

Superdense coding

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

$$|\Phi^-\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$$

$$|\Psi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}$$

$$|\Psi^-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

- Question: compute the following.

$$I_1 |\Phi^+\rangle = |\Phi^+\rangle$$

$$X_1 |\Phi^+\rangle = |\Psi^+\rangle$$

$$iY_1 |\Phi^+\rangle = |\Psi^-\rangle$$

$$Z_1 |\Phi^+\rangle = |\Phi^-\rangle$$

These are unitary and can be implemented w/ *quantum gates*.

Superdense coding

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

$$|\Phi^-\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$$

$$|\Psi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}$$

$$|\Psi^-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

- On the other hand, show that all the Bell states are orthonormal (Ex. 2.69)
- What do we know about orthonormal states?
  - There's a measurement distinguishing them.
  - Exercise: write down a unitary matrix *M* sending the 1st one to 00, 2nd to 01, 3rd to 10, 4th to 11. Express w/ quantum gates (later).

Superdense coding

$$00: \quad |\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$
$$01: \quad |\Phi^-\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$$
$$10: \quad |\Psi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}$$
$$11: \quad |\Psi^-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

Putting 2 and 2 together, this suggests the following communication scheme:

1. Prepare 2 qbits in state $|\Phi^+\rangle$
2. Send one to Romeo, one to Juliet
3. Romeo transforms the state according to the message he wants to send, using I_1, Z_1, X_1, iY_1 respectively
   - Only has to operate on his own qbit
4. Romeo sends his qbit back to Juliet
5. Juliet performs a measurement, obtaining Romeo's message from (3).

(4) takes time, so no superluminal flirtation. But only one qbit has to be sent -> superdensity.

Superdense coding

Question: is it secure? (Ex. 2.70)

- We mean: suppose Tybalt intercepts the qbit in transit in step (4).
- But his measurement operators have the form $M_i = M'_i \otimes I$
- Question: compute (2 ways) the measurement probabilities applied to the Bell states.
    - Can't distinguish them. Secure.

$$00: \quad |\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

$$01: \quad |\Phi^-\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}$$

$$10: \quad |\Psi^+\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}$$

$$11: \quad |\Psi^-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

1. Prepare 2 qbits in state 00 :
2. Send one to Romeo, one to Juliet
3. Romeo transforms the state according to the message he wants to send
4. Romeo sends his qbit back to Juliet
5. Juliet performs a measurement, obtaining Romeo's message from (3).