

Quantum Mechanics Overview

Many people like to be surprised by a trip. This seems to be the perspective of the authors of most of the popular undergraduate quantum mechanics (QM) texts. They pick a place (often very different places), and just begin the tour. I prefer to study the map before the tour and have a good idea of where I'm going. This document is intended to provide that map before we jump in with our authors. Hopefully this approach will provide a happy medium between studying the map in advance and being surprised. There will still be plenty of surprises, but this will give some sense of the big picture. It may not all make sense right away. That's ok. You can come back to it as your journey into the subject unfolds.

When we studied classical mechanics (CM), our picture of what we were trying to do could be schematically shown as Fig 1. (When I presented this figure at the beginning of the CM notes, the labels were position and velocity. Now that we have generalized coordinates and momenta, we can generalize the labels.) The whole game in CM was to take the initial state of a system as described by the initial values of the generalized coordinates and generalized velocities (or momenta) and find a way to evolve that state forward in time so as to make a prediction.

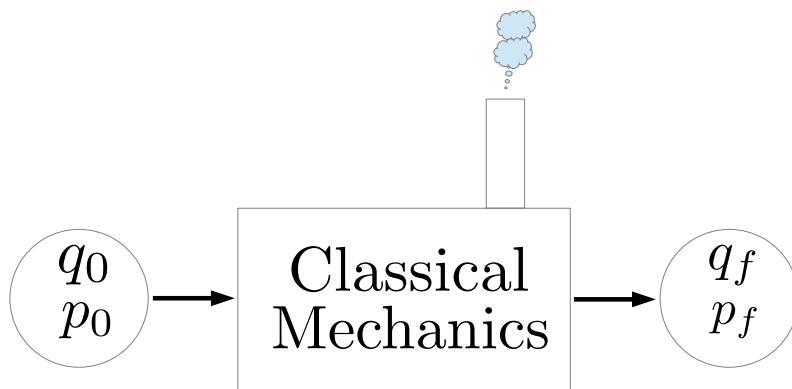


Figure 1: CM schematic.

CM is a perfectly fine theory (consistent, predictive, etc.). However, its predictions do not always match experiment, as you are no doubt aware from A&N. Hence we need a new theory. Ideally that theory should contain a limit that matches CM when we deal with macroscopic bodies (those containing moles of fundamental particles), after all, CM gets the right answers there! The situation we find ourselves in is logically similar to the development of special relativity. There, we needed a new theory that reduced to nonrelativistic mechanics when the speeds are slow, but also gets the right answers as we approach the speed of light. However to do so, we had to radically rethink our basic assumptions. As you already know from A&N, building QM requires a similar rethinking.

There is another analogy with CM that is useful to point out. In introductory physics, you get some basic experience with CM. In 231, we made that experience more systematic, learned more sophisticated techniques, and expanded the kinds of systems you could tackle. If you took or take another mechanics class, that experience will repeat. This is a pattern in physics. Repeated exposures to electricity & magnetism and QM follow this same pattern. So here, you'll meet many of the same ideas that you met in A&N, but we will make them more systematic, technically sophisticated, and general.

Moving back to the structure of QM, the development of any theory is initiated with a *need* for a new theory that comes in the form of experimental observations that cannot be predicted

with the old theory (experimental problems) and/or logical inconsistencies between parts of the existing theory (theory problems). As a new theory takes shape, it must then solve the original problems, be a complete logical structure in its own right, and make new predictions beyond solving the initial problems that prompted its development. QM took its current form via a complicated iterative process through these steps that we will not follow here. Some of this history is laid out in the A&N text and elsewhere. My goal here is to describe the subject as it is today.

QM is not a slightly tweaked way of doing CM. The starting assumptions and the goals are completely different¹ (or at least very different)! We cannot specify the initial state of the system with perfectly known initial conditions for all of the generalized coordinates and momenta. We now understand this as a consequence of the uncertainty principle. We also can't exactly predict the final state of the system, but only a probability distribution for what the system may do. So our theory will involve a completely new set of constructs for specifying the initial and final states of the system, and our goal will be to generate a probability distribution for what may happen in an experiment instead of exactly how it will turn out. That said, our experience with CM is still quite relevant to developing QM. The Lagrangian and Hamiltonian formalisms play a starring role in quantum theory. And since humans knew CM before QM, there is a systematic way of turning a classical theory into a quantum one that is called Canonical Quantization. This basically adds the assumption of uncertainty relations (like Heisenberg $\Delta x \Delta p \leq \hbar/2$, but there are more). The existence of these quantization rules are the foundational new assumption of QM and they initiate the development of the mathematical and logical structure that I describe next.

Since our predictions will be probabilistic, one of our “new” constructs for describing nature will be the familiar notions of probability. I’m attaching some probability review from Ref. [1] to the end of this document as Appendix A. It’s optional reading for now. Read it now if you like or come back to it as needed later. However probability alone (which is really the thing we observe) is insufficient to build the theory. Instead we must invent a new notion called probability amplitude, which is a complex number whose absolute value squared is the probability. As a part of today’s reading, read Appendix B from Stewart’s calculus text as a way of expanding and/or refreshing your knowledge of complex numbers. We’ll find that each outcome of our experiments has a probability and each way that this outcome could occur has a probability amplitude. So our goals in QM will consist of developing the rules for working with these probability amplitudes to calculate the probability of some observation. We’ll need to:

1. Find a way to specify the state of a system since exact values of position, momentum, angular momentum, etc are not going to work. The solution that has been found is another new construct. We will represent the state of the system using a vector in a new kind of vector space, called Hilbert space, which is separate from the real, 4D space in which we live. These state vectors, or Hilbert space vectors are written with something called Dirac notation, as things like this $|\text{variable they describe}\rangle$, eg $|+z\rangle$, for an electron with spin oriented along the $+z$ axis. Since I’ve now used the word “vector space” you might imagine that we’re going to apply some ideas from linear algebra here, and you’d be right. Hence I’m attaching as Appendix C some linear algebra review from Ref. [1] to the end of this overview which you should also read as a part of today’s reading.
2. The dot products of one of these Hilbert space vectors describing the current state of the system with a Hilbert space vector describing a possible observed state of the system provides a probability amplitude for a possible observation. So the amplitude to find a

¹There is historical precedent for this. Aristotelian “physicists” sought to explain the existence of motion. Newtonian physics reframes the question as explaining acceleration rather than motion.

particle that we already know is in the $|+z\rangle$ state in the $| -z \rangle$ state is given by the dot product $\langle -z | +z \rangle = 0$ using our funny Dirac notation for the dot product. In this case the amplitude is just zero because if we know it's in the $|+z\rangle$ state then its amplitude, and thus its probability, to be in the $| -z \rangle$ state is zero.

3. There are rules for how to combine the amplitudes if an observed outcome could occur by more than one path or in multiple steps. Once we have the amplitude, we can take its absolute square to get the probability.
4. The probability of an event may evolve in time. We'll want to calculate this time evolution of the probability and to do so, it's underlying constructs of Hilbert space vectors and amplitudes.

Note that it's really the last step that is most analogous to the classical mechanics machine. In CM, the way to describe a state of a system (at least a simple one) is pretty intuitive. In QM, we'll have a lot of work to do just to build the technology for describing a state before we can even begin to worry about how it will evolve in time.

Does all of this sound like the QM you know from A&N? Maybe not. In A&N you mostly learned wave mechanics, which is a part of a version of the above picture that's useful when the question at hand is related to the position and/or momentum of the particle. We'll cycle back to wave mechanics in Townsend's chapter 6 [3], but let me see if I can roughly re-cast your experience in wave mechanics using the broader language above. You learned about the wavefunction $\psi(x)$. This is an example of an amplitude, or more accurately an amplitude density to find a particle at a position x . You might recall that you computed the probability $P(x, x + \Delta x)$ of finding the particle in some region between x and $x + \Delta x$ as follows:

$$P(x, x + \Delta x) = \int_x^{x+\Delta x} |\psi(x)|^2 dx. \quad (1)$$

Since the wave function is an amplitude density, squaring it gives a probability density (probability per length). So integrating over some small region gives the probability of finding it in some window. All of that is an implementation of what I describe in point 3 above. The Schrödinger equation is an implementation of point 4 above, evolving these systems in time. In your A&N treatment, it works to sweep discussion of points 1 and 2 above about vectors in Hilbert space under the rug. This formalism is still present under the hood, but for the cases you studied, it's not necessary to make it explicit, but I can. Let me call the state of the particle $|\phi\rangle$. The amplitude to find the particle at a position x , which I already said was the wave function is then $\psi(x) = \langle x | \phi \rangle$.

This business of Hilbert space vectors was swept under the rug because the relevant vector space is an infinite dimensional one and that's a difficult place to begin a discussion of complex Hilbert spaces. It's of course much easier to begin thinking about these new kinds of vector spaces with questions that can be addressed with small, 2 and 3 dimensional Hilbert spaces before going to infinite dimensional ones. We'll return to wave mechanics in this course and see how the whole thing fits into the broader structure of QM after we get some experiences with lower-dimensional Hilbert spaces.

You might also be asking yourself at this stage, "where do these crazy new ideas, like amplitudes and Hilbert space vectors come from?" There are probably interesting historical stories for that, and there are ways to understand why they are a good choice. But in short, when a new theory is built, we cast around for the mathematical constructs we need to build a consistent, predictive theory that works. In this case, it's probability amplitudes, Hilbert space vectors, etc. that do the job. Newton invented calculus so that he could build a theory of mechanics. It

was found that Riemannian geometry is a framework for General Relativity. Vector calculus in three dimensions is the language of classical electricity and magnetism. So you shouldn't get tied in knots about the question, "why Hilbert space vectors?" any more than you should about asking, "why force diagrams?"

As a final note on the structure, I have a schematic for how QM works that is analogous to Fig. 1, but I think it's best to save it until we see a few systems. At that point, we can walk through it with an example in mind to which we can attach each step.

So how should we begin to explore QM? Or maybe first, why should we explore QM? A number of years ago the faculty and students in the department decided that a full term of 300-level QM should be required for physics majors. They felt that the prominent role played by QM in physics today makes this experience a part of the definition of being a physicist. Hence our study of QM should provide a sort of broad training to meet this department goal. Several of our students have now gone directly into the workforce in fields like quantum computing. There is evidence that these sorts of jobs are growing. Google "quantum workforce" and read articles from sources like *Forbes*. Other students go on to take a graduate QM course. Hence this course provides a practically useful background for students headed in a variety of directions.

With those reasons in mind, how should we start and what should we do? While subjects like E&M and classical mechanics have a fairly standard order, the physics community does not agree on how to organize a QM course. As far as I can tell, something like half of undergraduate programs use Griffiths [2] (earlier editions were just Griffiths, now it's Griffiths and Schroeter). Griffiths has a writing style that works well for modern undergraduates. However he starts with wave mechanics again and does not really, in my opinion, fully put together the pieces of the full framework described above. Roughly, something like the other half of undergraduate courses use Townsend [3]. Townsend begins with a 2D Hilbert space (rather than the infinite dimensional one of wave mechanics) and builds the full picture from there. This provides a way to develop the full picture in about the friendliest possible way. This approach matches what I think is the most popular graduate text by Sakurai [4], and several people have told me that Townsend provided a smoother transition to graduate QM. However, I personally find Townsend's writing style somewhat less friendly and conceptually rich relative to Griffiths. The approach used by Townsend and Sakurai, appears to have originated with the Feynman Lectures [5]. Feynman *et al.* contains more of the conversation that I find appealing. He also starts back a little further than Townsend, perhaps providing a better bridge from A&N. However, there are also problems with Feynman. There may be more distracting extra information, the treatment is slightly dated in several ways, and it isn't really a full traditional textbook, lacking a large number of problems and worked examples. As you can see, there is no perfect solution, and in fact a perfect solution is probably impossible. I've settled on Townsend with some supplementation from Feynman *et al.* and from me. We'll start with the friendly introduction of Feynman *et al.*, then pick up with Townsend when Feynman *et al.* gets to the place where Townsend starts (see the syllabus). This seems to me a nice way to balance the considerations above. There is another book by Beck [1] that follows a logical flow similar to Townsend and Feynman *et al.*. It's nice too (in addition to acknowledging a wonderful Carleton physics professor), but I think there is more optics than makes sense for our course here. I'll put Beck, along with the other texts cited here, plus some other classics, on the course reserve shelf if you want to read another treatment.

I have chosen our course texts for a combination of the content and clarity that they offer. My choices should not be understood as any special endorsement of the authors. This is particularly worthy of mention in the case of Feynman, with whom the field has a complicated relationship. Unfortunately the Feynman Lectures (from 1964) also exclusively use male pronouns. I seriously considered rejecting it based on this flaw, in spite of my strong motivation to use it based the quality of the physics. Ultimately, after extensive discussion with colleagues, I decided that it

doesn't make sense to try to hide the sexist past (and unfortunately present) of our field, but instead use these instances as a chance to reflect on the change that is needed going forward.

As a final note, the theory we'll build here should perhaps be referred to as nonrelativistic QM. To finish putting together a full theory of modern physics requires both relativity and QM. That combination results in relativistic quantum field theory, which requires yet another rethink of the basics, in this case, the notion of a particle. Perhaps near the end of the course, I'll offer some insight into what's need for that revolution.

References

- [1] M. Beck, *Quantum mechanics: theory and experiment*, Oxford University Press, New York, New York, (2012).
- [2] D.J. Griffiths and D.F. Schroeter, *Introduction to quantum mechanics*, 3rd Edition, Cambridge University Press, Cambridge, United Kingdom (2018).
- [3] J. Townsend, *A modern approach to quantum mechanics*, 2nd Edition, University Science Books, Mill Valley, California (2012).
- [4] J.J. Sakurai, *Modern quantum mechanics*, Addison-Wesley, Reading, Massachusetts (1994).
- [5] R. Feynman, R. Leighton, and M. Sands, *The Feynman lectures on physics*, Volume III, California Institute of Technology, (1965). Available in print and online https://www.feynmanlectures.caltech.edu/III_toc.html.

Appendix A: Probability

CHAPTER 1

Mathematical Preliminaries

Before beginning a discussion of quantum mechanics, it's useful to review some topics from mathematics and classical physics. In this chapter we'll go over some areas of probability theory and linear algebra that we'll find useful later. You may be familiar with the concepts presented here, but I suggest that you read through this chapter, if only to familiarize yourself with the notation and terminology we'll be using.

1.1 PROBABILITY AND STATISTICS

1.1.1 Moments of Measured Data

We want to measure a quantity x , which is a property of some object. So, we break out our x -meter and make measurements, getting N values of x : x_{m1} , x_{m2} , etc. We're using the subscript m here to indicate that these are measured values. Collectively we'll refer to these measurements as x_{mi} , where $i = 1, 2, \dots, N$.

We can calculate the average of x , $\langle x \rangle$, by adding up the measured values and dividing by the number of measurements:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^N x_{mi}. \quad (1.1)$$

We'll use the bracket symbol $\langle \dots \rangle$ to denote an average; it represents the average of whatever is inside the brackets. We'll also refer to $\langle x \rangle$ as the mean of x . We can calculate the average of x^n , $\langle x^n \rangle$, by simply using the definition of the average

$$\langle x^n \rangle = \frac{1}{N} \sum_{i=1}^N (x_{mi})^n. \quad (1.2)$$

Another name for $\langle x^n \rangle$ n^{th} -order moment of x .

Remember, the brackets $\langle \dots \rangle$ mean that we average *whatever quantity* is inside the brackets. We calculate this average by adding up the values and dividing by the number of measurements. The average of a function of x , $f(x)$, is thus

$$\langle f(x) \rangle = \frac{1}{N} \sum_{i=1}^N f(x_{mi}). \quad (1.3)$$

The average is linear. This means that

$$\langle A_1 f_1(x) + A_2 f_2(x) \rangle = A_1 \langle f_1(x) \rangle + A_2 \langle f_2(x) \rangle, \quad (1.4)$$

Where A_1 and A_2 are constants.

Frequently, we're interested in how far a typical measured value might be from the mean value. Let's define the deviation from the mean as

$$\delta x = x - \langle x \rangle. \quad (1.5)$$

The average of this quantity is 0, so it is not a good measure of the fluctuations:

$$\langle \delta x \rangle = \langle x - \langle x \rangle \rangle = \langle x \rangle - \langle \langle x \rangle \rangle = \langle x \rangle - \langle x \rangle = 0. \quad (1.6)$$

Here we have used both the linearity property [eq. (1.4)] and the fact that once a quantity has been averaged it becomes a constant. The average of a constant is the constant. $\langle \delta x \rangle$ is not a good measure of the fluctuations, because roughly half the time it's positive, and half the time it's negative, so it averages zero. To overcome this problem, we can average the square of δx . Calculating this, we see that

$$\begin{aligned} \langle (\delta x)^2 \rangle &= \langle (x - \langle x \rangle)^2 \rangle \\ &= \langle (x^2 - 2x\langle x \rangle + \langle x \rangle^2) \rangle \\ &= \langle x^2 \rangle + \langle (-2\langle x \rangle x) \rangle + \langle (\langle x \rangle^2) \rangle \quad (1.7) \\ &= \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2 \\ &= \langle x^2 \rangle - \langle x \rangle^2. \end{aligned}$$

The variance of x , Δx^2 , is defined as

$$\Delta x^2 \equiv \langle (\delta x)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2. \quad (1.8)$$

The standard deviation of x , Δx , is the square root of the variance:

$$\Delta x = \sqrt{\Delta x^2} = \left(\langle x^2 \rangle - \langle x \rangle^2 \right)^{1/2}. \quad (1.9)$$

The standard deviation is a measure of how far a particular measurement is likely to be from the mean, and we will use it to quantify the uncertainty in a series of measurements.¹

EXAMPLE 1.1

Measurements of x yield the values

$$x_m = 9, 5, 25, 23, 10, 22, 8, 8, 21, 20 \quad i = 1, \dots, 10. \quad (1.10)$$

Calculate the mean and standard deviation of this data.

To calculate the mean of x we use eq. (1.1):

$$\langle x \rangle = \frac{1}{10} (9 + 5 + 25 + 23 + 10 + 22 + 8 + 8 + 21 + 20) = 15.1. \quad (1.11)$$

In order to calculate the standard deviation, we first need to calculate $\langle x^2 \rangle$. Using eq. (1.2) we find

$$\langle x^2 \rangle = \frac{1}{10} (9^2 + 5^2 + 25^2 + 23^2 + 10^2 + 22^2 + 8^2 + 8^2 + 21^2 + 20^2) = 281.3. \quad (1.12)$$

Substituting these results into eq. (1.9) yields

$$\Delta x = \left[281.3 - (15.1)^2 \right]^{1/2} = 7.3. \quad (1.13)$$

Using the standard deviation as the uncertainty of the measurements, we say that $x = 15.1 \pm 7.3$.

1.1.2 Probability

Let's look at the data from example 1.1 [eq. (1.10)] a little differently, by creating a histogram of the data. We break the full range of the data into M segments, known as bins. The bins have equal widths, and we'll label them as x_j ($j = 1, 2, \dots, M$). Note that there's no subscript m on x_j , because we're labeling an x value corresponding to a bin (which is a value that *could* have been measured), not a particular measured value. We sort individual measurements into bins, then count the number of measurements in each bin, $N(x_j)$. This process, known as histogramming, is best illustrated with an example.

¹ Frequently the variance will be defined with a normalization factor of $1/(N-1)$ in front of the sum in eq. (1.3) instead of $1/N$, which makes the normalization factor for the standard deviation $1/\sqrt{N-1}$. For details on why this is the case, see secs. 8.1 and 10.2 of ref. [1.1]. The choice of $1/\sqrt{N-1}$ is common in calculators and spreadsheets. In the limit of large N the difference is insignificant. In this book we'll assume a $1/N$ normalization.

EXAMPLE 1.2

Create a histogram of the data from example 1.1.

The data values range from 5 to 25; let's break this interval into 7 equal-width bins. The first bin contains measurements which yield x values of 5, 6, and 7, and we'll label it bin $x_1 = 6$, corresponding to the center value. Bin $x_2 = 9$ contains values 8, 9 and 10, and so on.

Examining the data in eq. (1.10), we find that one measurement yields a value of 5, 6, or 7. Thus, the number of measurements in bin $x_1 = 6$ is $N(6) = 1$. Four of our measurements fall in the second bin (holding values 8, 9, and 10) so $N(9) = 4$. Continuing on in this vein, we can find the rest of the histogram values. A plot of the histogram is shown in fig. 1.1.

Histograms allow us to estimate the probability that we will obtain a particular measurement. The probability $P(x_j)$ that a measurement will fall into a particular bin x_j is simply the ratio of the number of measurements in that bin to the total number of measurements. In other words

$$P(x_j) = \frac{N(x_j)}{N}. \quad (1.14)$$

The probability distribution corresponding to the histogram calculated in example 1.2 is plotted in fig. 1.1. As can be seen, the probability distribution is a scaled version of the histogram.

Summing all the histogram values must yield the total number of measurements N . If the histogram has M bins then

$$\sum_{j=1}^M P(x_j) = \frac{\sum_{j=1}^M N(x_j)}{N} = \frac{N}{N} = 1. \quad (1.15)$$

Our probability is properly normalized, which means that the sum of the probabilities is 1.

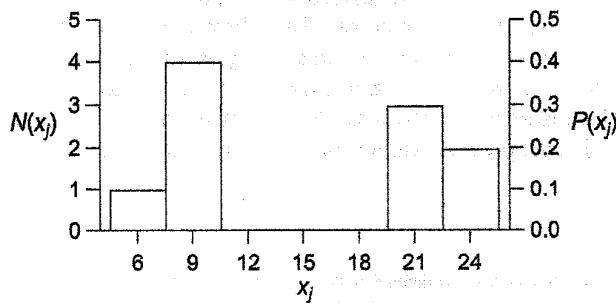


Fig 1.1 The histogram calculated in example 1.2 is plotted on the left axis, while the corresponding probability distribution is plotted on the right axis.

Equation (1.14) is only an estimate of the probability, and it's a fairly coarse estimate at that if the number of measurements N is small. Furthermore, with small N it is necessary to use fairly wide bins, so the x resolution of $P(x_j)$ is not very good. Conversely, if N is large, it is possible to obtain a fairly accurate and high-resolution estimate of $P(x_j)$.

One thing we can do with a probability distribution is simply plot it. This allows us to visualize our data, and see which measured values are likely to occur, and which are not very likely. For example, fig. 1.1 tells us that measurements of x are clumped into regions, which is more information than simply saying that $x = 15.1 \pm 7.3$ (from example 1.1). We can also use the probabilities to directly calculate moments, without having to go back to the original data. Given the probabilities $P(x_j)$ for $j = 1, 2, \dots, M$, the mean of x is given by

$$\langle x \rangle = \sum_{j=1}^M x_j P(x_j). \quad (1.16)$$

Thus, to calculate the mean, weight the value by its probability, and then sum over all possible values. In general the mean of a function of x , $f(x)$, is given by

$$\langle f(x) \rangle = \sum_{j=1}^M f(x_j) P(x_j). \quad (1.17)$$

EXAMPLE 1.3

Calculate the mean and standard deviation of the data from example 1.1, using the corresponding probability distribution.

Figure 1.1 shows the histogram of the data from example 1.1. We can use this histogram to estimate the probability distribution using eq. (1.14). This probability distribution is plotted on the right axis of fig. 1.1. Using these probabilities, eq. (1.16) tells us that

$$\langle x \rangle = (6)(0.1) + (9)(0.4) + (21)(0.3) + (24)(0.2) = 15.3, \quad (1.18)$$

and eq. (1.17) says that

$$\langle x^2 \rangle = (6)^2(0.1) + (9)^2(0.4) + (21)^2(0.3) + (24)^2(0.2) = 283.5. \quad (1.19)$$

The standard deviation is then

$$\Delta x = \left[283.5 - (15.3)^2 \right]^{1/2} = 7.0. \quad (1.20)$$

Note that the calculation of $\langle x \rangle$ using the estimated probability distribution in example 1.3 is not in perfect agreement with the direct calculation in example 1.1, although the two determinations of $\langle x \rangle$ agree to well within one standard deviation. The agreement is not perfect because the probability distribution in example 1.3 is just an estimate of the true distribution. In the limit that the number of measurements is very large, and the resolution of $P(x_j)$ is very fine, the estimated distribution will approach the true distribution, and the

mean calculated using the probabilities will approach the mean determined directly from the data.

1.1.3 Continuous Probability Distributions

So far we've talked about probabilities that are determined at discrete values of x , for example, $P(x_j)$ for $j = 1, 2, \dots, M$. When talking about real data we always have a finite measurement resolution, so measured probability distributions will always be discrete. Theoretically, however, it is possible to discuss probability distributions of a continuous variable.

Consider the position of a particle x . Position is a continuous variable and, in principle, the particle can be anywhere. Since it can be *anywhere*, the probability that it will be in any *particular* place is zero. For example, the probability of finding the particle at *exactly* $x = 9.999\dots$ is zero—it doesn't make sense to talk about it. What it does make sense to talk about is the probability of finding the particle within some range of positions, say between $x = 9.99$ and $x = 10.00$. If the range is small, the probability of finding the particle will be proportional to the range (e.g., a particle is twice as likely to be found within a $2\text{ }\mu\text{m}$ interval than within a $1\text{ }\mu\text{m}$ interval). If we take dx to be a small-length interval (in the sense of a differential), then the probability $P(x)$ that the particle will be found between x and $x+dx$ is

$$P(x) = p(x)dx, \quad (1.21)$$

where $p(x)$ is called the probability density of x .² Continuous distributions are normalized by integrating the probability density over its entire range. If a particle can be found anywhere between $x = -\infty$ and $x = \infty$, the normalization condition is

$$\int_{-\infty}^{\infty} p(x)dx = 1. \quad (1.22)$$

For discrete probabilities we calculate the mean by weighting a value by its probability, and summing over all possible values. For continuous distributions we do the same thing, except that the sum is replaced by an integral. We thus have

$$\langle x \rangle = \int_{-\infty}^{\infty} xp(x)dx, \quad (1.23)$$

and

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x)p(x)dx. \quad (1.24)$$

If we wish to calculate the standard deviation of x , we do it the same way as we did in the discrete case, using eq. (1.9).

2. The probability on the left-hand side of eq. (1.21) is dimensionless, while the interval dx has units of length (m). For the units in eq. (1.21) to work out, then $p(x)$ must have units of inverse length (m^{-1}). A probability density always has the inverse units of its argument.

1.1.4 Joint and Conditional Probabilities

Sometimes we're interested in the probability of more than one thing. For example, we want to know the probability that a particle has a particular value of x *and* a particular value of y (i.e., it has both of these values simultaneously). We denote this probability by $P(x, y)$, and refer to it as the joint probability of x and y .

The joint probability density $p(x, y)$, has the property that integrating with respect to one variable yields the probability density of the other. For example,

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy. \quad (1.25)$$

Here we say that $p(x)$ is the marginal probability density of x . Averages are obtained by integrating with respect to both variables:

$$\langle f(x, y) \rangle = \int_{-\infty}^{\infty} f(x, y) p(x, y) dx dy. \quad (1.26)$$

In some situations we want to know the probability of obtaining x , given a particular value of y . We denote this probability by $P(x|y)$, and call it a conditional probability, because it represents the probability of x conditioned on y . The joint and conditional probabilities are related by

$$P(x, y) = P(x|y) P(y). \quad (1.27)$$

For more details see ref. [1.1], sec. 3.7.

1.2 LINEAR ALGEBRA

1.2.1 Vectors and Basis Sets

We'll use what you already know about vectors from first-year physics to develop a vocabulary for linear algebra, and then we'll build on that.

We can express a vector \mathbf{a} (we'll denote vectors with bold type), that "lives" in the x - y plane as

$$\mathbf{a} = a_x \mathbf{u}_x + a_y \mathbf{u}_y, \quad (1.28)$$

where \mathbf{u}_x and \mathbf{u}_y are dimensionless unit vectors that point in the positive x - and y -directions respectively.³ Any vector in the x - y plane can be expressed by giving its components (coordinates) in the \mathbf{u}_x and \mathbf{u}_y directions; in eq. (1.28) the components are a_x and a_y . The vectors \mathbf{u}_x and \mathbf{u}_y are called basis vectors, and they make up a basis set. This basis set contains two vectors, because it is a 2-D vector space (an N -dimensional vector space must have N basis vectors).

³ We are not using the more familiar notion of \hat{i} and \hat{j} , or \hat{x} and \hat{y} , to denote the unit vectors because we will reserve the caret symbol $\hat{\cdot}$ for something else.

Appendix B: Complex Numbers

COMPLEX NUMBERS

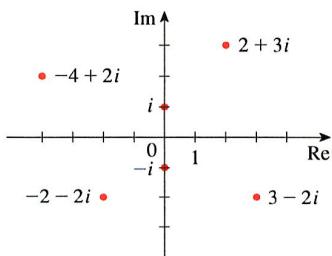


FIGURE 1

Complex numbers as points in the Argand plane

A **complex number** can be represented by an expression of the form $a + bi$, where a and b are real numbers and i is a symbol with the property that $i^2 = -1$. The complex number $a + bi$ can also be represented by the ordered pair (a, b) and plotted as a point in a plane (called the Argand plane) as in Figure 1. Thus, the complex number $i = 0 + 1 \cdot i$ is identified with the point $(0, 1)$.

The **real part** of the complex number $a + bi$ is the real number a and the **imaginary part** is the real number b . Thus, the real part of $4 - 3i$ is 4 and the imaginary part is -3 . Two complex numbers $a + bi$ and $c + di$ are **equal** if $a = c$ and $b = d$, that is, their real parts are equal and their imaginary parts are equal. In the Argand plane the horizontal axis is called the **real axis** and the vertical axis is called the **imaginary axis**.

The sum and difference of two complex numbers are defined by adding or subtracting their real parts and their imaginary parts:

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

$$(a + bi) - (c + di) = (a - c) + (b - d)i$$

For instance,

$$(1 - i) + (4 + 7i) = (1 + 4) + (-1 + 7)i = 5 + 6i$$

The product of complex numbers is defined so that the usual commutative and distributive laws hold:

$$\begin{aligned} (a + bi)(c + di) &= a(c + di) + (bi)(c + di) \\ &= ac + adi + bci + bdi^2 \end{aligned}$$

Since $i^2 = -1$, this becomes

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

EXAMPLE 1

$$\begin{aligned} (-1 + 3i)(2 - 5i) &= (-1)(2 - 5i) + 3i(2 - 5i) \\ &= -2 + 5i + 6i - 15(-1) = 13 + 11i \end{aligned}$$

others use a * for this

Division of complex numbers is much like rationalizing the denominator of a rational expression. For the complex number $z = a + bi$, we define its **complex conjugate** to be $\bar{z} = a - bi$. To find the quotient of two complex numbers we multiply numerator and denominator by the complex conjugate of the denominator.

EXAMPLE 2 Express the number $\frac{-1 + 3i}{2 + 5i}$ in the form $a + bi$.

SOLUTION We multiply numerator and denominator by the complex conjugate of $2 + 5i$, namely $2 - 5i$, and we take advantage of the result of Example 1:

$$\frac{-1 + 3i}{2 + 5i} = \frac{-1 + 3i}{2 + 5i} \cdot \frac{2 - 5i}{2 - 5i} = \frac{13 + 11i}{2^2 + 5^2} = \frac{13}{29} + \frac{11}{29}i$$

The geometric interpretation of the complex conjugate is shown in Figure 2: \bar{z} is the reflection of z in the real axis. We list some of the properties of the complex conjugate in the following box. The proofs follow from the definition and are requested in Exercise 18.

Properties of Conjugates

$$\overline{z + w} = \bar{z} + \bar{w} \quad \overline{zw} = \bar{z} \bar{w} \quad \overline{z^n} = \bar{z}^n$$

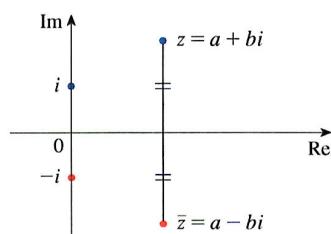


FIGURE 2

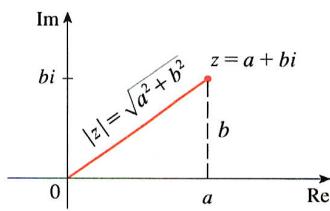


FIGURE 3

The **modulus**, or **absolute value**, $|z|$ of a complex number $z = a + bi$ is its distance from the origin. From Figure 3 we see that if $z = a + bi$, then

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

Notice that

$$z\bar{z} = (a + bi)(a - bi) = a^2 + abi - abi - b^2i^2 = a^2 + b^2$$

and so

$$z\bar{z} = |z|^2$$

This explains why the division procedure in Example 2 works in general:

$$\frac{z}{w} = \frac{z\bar{w}}{w\bar{w}} = \frac{z\bar{w}}{|w|^2}$$

Since $i^2 = -1$, we can think of i as a square root of -1 . But notice that we also have $(-i)^2 = i^2 = -1$ and so $-i$ is also a square root of -1 . We say that i is the **principal square root** of -1 and write $\sqrt{-1} = i$. In general, if c is any positive number, we write $\sqrt{-c} = \sqrt{c} i$

With this convention, the usual derivation and formula for the roots of the quadratic equation $ax^2 + bx + c = 0$ are valid even when $b^2 - 4ac < 0$:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

EXAMPLE 3 Find the roots of the equation $x^2 + x + 1 = 0$.

SOLUTION Using the quadratic formula, we have

$$x = \frac{-1 \pm \sqrt{1^2 - 4 \cdot 1}}{2} = \frac{-1 \pm \sqrt{-3}}{2} = \frac{-1 \pm \sqrt{3}i}{2}$$

We observe that the solutions of the equation in Example 3 are complex conjugates of each other. In general, the solutions of any quadratic equation $ax^2 + bx + c = 0$ with real coefficients a , b , and c are always complex conjugates. (If z is real, $\bar{z} = z$, so z is its own conjugate.)

We have seen that if we allow complex numbers as solutions, then every quadratic equation has a solution. More generally, it is true that every polynomial equation

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

of degree at least one has a solution among the complex numbers. This fact is known as the Fundamental Theorem of Algebra and was proved by Gauss.

POLAR FORM

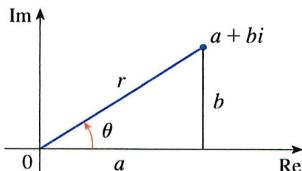


FIGURE 4

We know that any complex number $z = a + bi$ can be considered as a point (a, b) and that any such point can be represented by polar coordinates (r, θ) with $r \geq 0$. In fact,

$$a = r \cos \theta \quad b = r \sin \theta$$

as in Figure 4. Therefore, we have

$$z = a + bi = (r \cos \theta) + (r \sin \theta)i$$

Thus, we can write any complex number z in the form

$$z = r(\cos \theta + i \sin \theta)$$

where $r = |z| = \sqrt{a^2 + b^2}$ and $\tan \theta = \frac{b}{a}$

The angle θ is called the **argument** of z and we write $\theta = \arg(z)$. Note that $\arg(z)$ is not unique; any two arguments of z differ by an integer multiple of 2π .

EXAMPLE 4 Write the following numbers in polar form.

(a) $z = 1 + i$ (b) $w = \sqrt{3} - i$

SOLUTION

(a) We have $r = |z| = \sqrt{1^2 + 1^2} = \sqrt{2}$ and $\tan \theta = 1$, so we can take $\theta = \pi/4$. Therefore, the polar form is

$$z = \sqrt{2} \left(\cos \frac{\pi}{4} + i \sin \frac{\pi}{4} \right)$$

(b) Here we have $r = |w| = \sqrt{3 + 1} = 2$ and $\tan \theta = -1/\sqrt{3}$. Since w lies in the fourth quadrant, we take $\theta = -\pi/6$ and

$$w = 2 \left[\cos \left(-\frac{\pi}{6} \right) + i \sin \left(-\frac{\pi}{6} \right) \right]$$

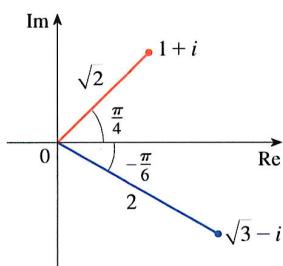


FIGURE 5

The numbers z and w are shown in Figure 5. ■

The polar form of complex numbers gives insight into multiplication and division. Let

$$z_1 = r_1(\cos \theta_1 + i \sin \theta_1) \quad z_2 = r_2(\cos \theta_2 + i \sin \theta_2)$$

be two complex numbers written in polar form. Then

$$\begin{aligned} z_1 z_2 &= r_1 r_2 (\cos \theta_1 + i \sin \theta_1)(\cos \theta_2 + i \sin \theta_2) \\ &= r_1 r_2 [(\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + i(\sin \theta_1 \cos \theta_2 + \cos \theta_1 \sin \theta_2)] \end{aligned}$$

Therefore, using the addition formulas for cosine and sine, we have

$$z_1 z_2 = r_1 r_2 [\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)]$$

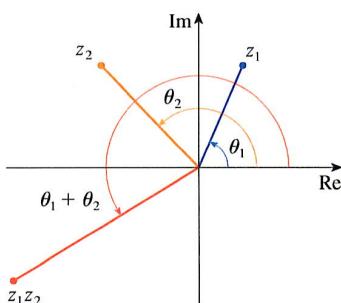


FIGURE 6

This formula says that *to multiply two complex numbers we multiply the moduli and add the arguments*. (See Figure 6.)

A similar argument using the subtraction formulas for sine and cosine shows that *to divide two complex numbers we divide the moduli and subtract the arguments*.

$$\frac{z_1}{z_2} = \frac{r_1}{r_2} [\cos(\theta_1 - \theta_2) + i \sin(\theta_1 - \theta_2)] \quad z_2 \neq 0$$

In particular, taking $z_1 = 1$ and $z_2 = z$, (and therefore $\theta_1 = 0$ and $\theta_2 = \theta$), we have the following, which is illustrated in Figure 7.

$$\text{If } z = r(\cos \theta + i \sin \theta), \text{ then } \frac{1}{z} = \frac{1}{r} (\cos \theta - i \sin \theta).$$

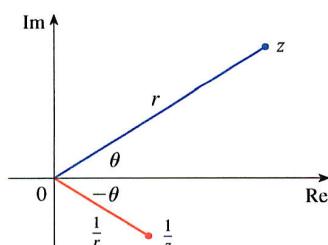


FIGURE 7

EXAMPLE 5 Find the product of the complex numbers $1 + i$ and $\sqrt{3} - i$ in polar form.

SOLUTION From Example 4 we have

$$1 + i = \sqrt{2} \left(\cos \frac{\pi}{4} + i \sin \frac{\pi}{4} \right)$$

$$\text{and } \sqrt{3} - i = 2 \left[\cos \left(-\frac{\pi}{6} \right) + i \sin \left(-\frac{\pi}{6} \right) \right]$$

So, by Equation 1,

$$\begin{aligned} (1 + i)(\sqrt{3} - i) &= 2\sqrt{2} \left[\cos \left(\frac{\pi}{4} - \frac{\pi}{6} \right) + i \sin \left(\frac{\pi}{4} - \frac{\pi}{6} \right) \right] \\ &= 2\sqrt{2} \left(\cos \frac{\pi}{12} + i \sin \frac{\pi}{12} \right) \end{aligned}$$

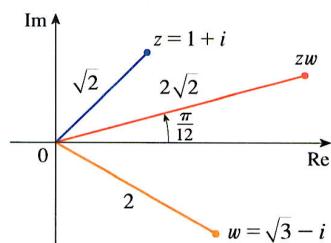


FIGURE 8

This is illustrated in Figure 8. ■

Repeated use of Formula 1 shows how to compute powers of a complex number. If

$$z = r(\cos \theta + i \sin \theta)$$

$$\text{then } z^2 = r^2(\cos 2\theta + i \sin 2\theta)$$

$$\text{and } z^3 = zz^2 = r^3(\cos 3\theta + i \sin 3\theta)$$

In general, we obtain the following result, which is named after the French mathematician Abraham De Moivre (1667–1754).

This part is
fun but
optional.

[2] De Moivre's Theorem If $z = r(\cos \theta + i \sin \theta)$ and n is a positive integer, then

$$z^n = [r(\cos \theta + i \sin \theta)]^n = r^n(\cos n\theta + i \sin n\theta)$$

This says that *to take the nth power of a complex number we take the nth power of the modulus and multiply the argument by n*.

EXAMPLE 6 Find $\left(\frac{1}{2} + \frac{1}{2}i\right)^{10}$.

SOLUTION Since $\frac{1}{2} + \frac{1}{2}i = \frac{1}{2}(1 + i)$, it follows from Example 4(a) that $\frac{1}{2} + \frac{1}{2}i$ has the polar form

$$\frac{1}{2} + \frac{1}{2}i = \frac{\sqrt{2}}{2} \left(\cos \frac{\pi}{4} + i \sin \frac{\pi}{4} \right)$$

So by De Moivre's Theorem,

$$\begin{aligned} \left(\frac{1}{2} + \frac{1}{2}i \right)^{10} &= \left(\frac{\sqrt{2}}{2} \right)^{10} \left(\cos \frac{10\pi}{4} + i \sin \frac{10\pi}{4} \right) \\ &= \frac{2^5}{2^{10}} \left(\cos \frac{5\pi}{2} + i \sin \frac{5\pi}{2} \right) = \frac{1}{32}i \end{aligned}$$

De Moivre's Theorem can also be used to find the n th roots of complex numbers. An n th root of the complex number z is a complex number w such that

$$w^n = z$$

Writing these two numbers in trigonometric form as

$$w = s(\cos \phi + i \sin \phi) \quad \text{and} \quad z = r(\cos \theta + i \sin \theta)$$

and using De Moivre's Theorem, we get

$$s^n(\cos n\phi + i \sin n\phi) = r(\cos \theta + i \sin \theta)$$

The equality of these two complex numbers shows that

$$s^n = r \quad \text{or} \quad s = r^{1/n}$$

$$\text{and} \quad \cos n\phi = \cos \theta \quad \text{and} \quad \sin n\phi = \sin \theta$$

opt. anal

From the fact that sine and cosine have period 2π it follows that

$$n\phi = \theta + 2k\pi \quad \text{or} \quad \phi = \frac{\theta + 2k\pi}{n}$$

$$\text{Thus} \quad w = r^{1/n} \left[\cos \left(\frac{\theta + 2k\pi}{n} \right) + i \sin \left(\frac{\theta + 2k\pi}{n} \right) \right]$$

Since this expression gives a different value of w for $k = 0, 1, 2, \dots, n - 1$, we have the following.

3 Roots of a Complex Number Let $z = r(\cos \theta + i \sin \theta)$ and let n be a positive integer. Then z has the n distinct n th roots

$$w_k = r^{1/n} \left[\cos \left(\frac{\theta + 2k\pi}{n} \right) + i \sin \left(\frac{\theta + 2k\pi}{n} \right) \right]$$

where $k = 0, 1, 2, \dots, n - 1$.

Notice that each of the n th roots of z has modulus $|w_k| = r^{1/n}$. Thus, all the n th roots of z lie on the circle of radius $r^{1/n}$ in the complex plane. Also, since the argument of each successive n th root exceeds the argument of the previous root by $2\pi/n$, we see that the n th roots of z are equally spaced on this circle.

EXAMPLE 7 Find the six sixth roots of $z = -8$ and graph these roots in the complex plane.

SOLUTION In trigonometric form, $z = 8(\cos \pi + i \sin \pi)$. Applying Equation 3 with $n = 6$, we get

$$w_k = 8^{1/6} \left(\cos \frac{\pi + 2k\pi}{6} + i \sin \frac{\pi + 2k\pi}{6} \right)$$

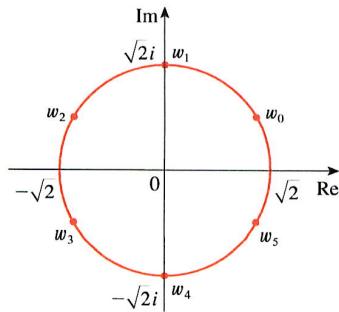
We get the six sixth roots of -8 by taking $k = 0, 1, 2, 3, 4, 5$ in this formula:

$$w_0 = 8^{1/6} \left(\cos \frac{\pi}{6} + i \sin \frac{\pi}{6} \right) = \sqrt{2} \left(\frac{\sqrt{3}}{2} + \frac{1}{2}i \right)$$

$$w_1 = 8^{1/6} \left(\cos \frac{\pi}{2} + i \sin \frac{\pi}{2} \right) = \sqrt{2}i$$

$$w_2 = 8^{1/6} \left(\cos \frac{5\pi}{6} + i \sin \frac{5\pi}{6} \right) = \sqrt{2} \left(-\frac{\sqrt{3}}{2} + \frac{1}{2}i \right)$$



**FIGURE 9**The six sixth roots of $z = -8$

$$w_3 = 8^{1/6} \left(\cos \frac{7\pi}{6} + i \sin \frac{7\pi}{6} \right) = \sqrt{2} \left(-\frac{\sqrt{3}}{2} - \frac{1}{2}i \right)$$

$$w_4 = 8^{1/6} \left(\cos \frac{3\pi}{2} + i \sin \frac{3\pi}{2} \right) = -\sqrt{2}i$$

$$w_5 = 8^{1/6} \left(\cos \frac{11\pi}{6} + i \sin \frac{11\pi}{6} \right) = \sqrt{2} \left(\frac{\sqrt{3}}{2} - \frac{1}{2}i \right)$$

All these points lie on the circle of radius $\sqrt{2}$ as shown in Figure 9. ■

end optional
COMPLEX EXPONENTIALS

We also need to give a meaning to the expression e^z when $z = x + iy$ is a complex number. The theory of infinite series as developed in Chapter 8 can be extended to the case where the terms are complex numbers. Using the Taylor series for e^x (8.7.12) as our guide, we define

$$\boxed{4} \quad e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots$$

and it turns out that this complex exponential function has the same properties as the real exponential function. In particular, it is true that

$$\boxed{5} \quad e^{z_1+z_2} = e^{z_1}e^{z_2}$$

If we put $z = iy$, where y is a real number, in Equation 4, and use the facts that

$$i^2 = -1, \quad i^3 = i^2i = -i, \quad i^4 = 1, \quad i^5 = i, \quad \dots$$

we get

$$\begin{aligned} e^{iy} &= 1 + iy + \frac{(iy)^2}{2!} + \frac{(iy)^3}{3!} + \frac{(iy)^4}{4!} + \frac{(iy)^5}{5!} + \dots \\ &= 1 + iy - \frac{y^2}{2!} - i \frac{y^3}{3!} + \frac{y^4}{4!} + i \frac{y^5}{5!} + \dots \\ &= \left(1 - \frac{y^2}{2!} + \frac{y^4}{4!} - \frac{y^6}{6!} + \dots \right) + i \left(y - \frac{y^3}{3!} + \frac{y^5}{5!} - \dots \right) \\ &= \cos y + i \sin y \end{aligned}$$

Here we have used the Taylor series for $\cos y$ and $\sin y$ (Equations 8.7.17 and 8.7.16). The result is a famous formula called **Euler's formula**:

$$\boxed{6} \quad e^{iy} = \cos y + i \sin y$$

Combining Euler's formula with Equation 5, we get

$$\boxed{7} \quad e^{x+iy} = e^x e^{iy} = e^x (\cos y + i \sin y)$$

We could write the result of Example 8(a) as

$$e^{i\pi} + 1 = 0$$

This equation relates the five most famous numbers in all of mathematics: 0, 1, e , i , and π .

EXAMPLE 8 Evaluate: (a) $e^{i\pi}$ (b) $e^{-1+i\pi/2}$

SOLUTION

(a) From Euler's equation (6) we have

$$e^{i\pi} = \cos \pi + i \sin \pi = -1 + i(0) = -1$$

(b) Using Equation 7 we get

$$e^{-1+i\pi/2} = e^{-1} \left(\cos \frac{\pi}{2} + i \sin \frac{\pi}{2} \right) = \frac{1}{e} [0 + i(1)] = \frac{i}{e}$$

Finally, we note that Euler's equation provides us with an easier method of proving De Moivre's Theorem:

$$[r(\cos \theta + i \sin \theta)]^n = (re^{i\theta})^n = r^n e^{in\theta} = r^n (\cos n\theta + i \sin n\theta)$$

EXERCISES

A Click here for answers.

S Click here for solutions.

1–14 Evaluate the expression and write your answer in the form $a + bi$.

1. $(5 - 6i) + (3 + 2i)$

3. $(2 + 5i)(4 - i)$

5. $\overline{12 + 7i}$

7. $\frac{1 + 4i}{3 + 2i}$

9. $\frac{1}{1 + i}$

11. i^3

13. $\sqrt{-25}$

2. $(4 - \frac{1}{2}i) - (9 + \frac{5}{2}i)$

4. $(1 - 2i)(8 - 3i)$

6. $\overline{2i(\frac{1}{2} - i)}$

8. $\frac{3 + 2i}{1 - 4i}$

10. $\frac{3}{4 - 3i}$

12. i^{100}

14. $\sqrt{-3}\sqrt{-12}$

15–17 Find the complex conjugate and the modulus of the number.

15. $12 - 5i$

16. $-1 + 2\sqrt{2}i$

17. $-4i$

18. Prove the following properties of complex numbers.

(a) $\bar{z} + w = \bar{z} + \bar{w}$ (b) $\bar{zw} = \bar{z}\bar{w}$

(c) $\bar{z}^n = \bar{z}^n$, where n is a positive integer

[Hint: Write $z = a + bi$, $w = c + di$.]

19–24 Find all solutions of the equation.

19. $4x^2 + 9 = 0$

20. $x^4 = 1$

21. $x^2 + 2x + 5 = 0$

22. $2x^2 - 2x + 1 = 0$

23. $z^2 + z + 2 = 0$

24. $z^2 + \frac{1}{2}z + \frac{1}{4} = 0$

25–28 Write the number in polar form with argument between 0 and 2π .

25. $-3 + 3i$

26. $1 - \sqrt{3}i$

27. $3 + 4i$

28. $8i$

29–32 Find polar forms for zw , z/w , and $1/z$ by first putting z and w into polar form.

29. $z = \sqrt{3} + i$, $w = 1 + \sqrt{3}i$

30. $z = 4\sqrt{3} - 4i$, $w = 8i$

31. $z = 2\sqrt{3} - 2i$, $w = -1 + i$

32. $z = 4(\sqrt{3} + i)$, $w = -3 - 3i$

33–36 Find the indicated power using De Moivre's Theorem.

33. $(1 + i)^{20}$

34. $(1 - \sqrt{3}i)^5$

35. $(2\sqrt{3} + 2i)^5$

36. $(1 - i)^8$

37–40 Find the indicated roots. Sketch the roots in the complex plane.

37. The eighth roots of 1

38. The fifth roots of 32

39. The cube roots of i

40. The cube roots of $1 + i$

41–46 Write the number in the form $a + bi$.

41. $e^{i\pi/2}$

42. $e^{2\pi i}$

43. $e^{i\pi/3}$

44. $e^{-i\pi}$

45. $e^{2+i\pi}$

46. $e^{\pi+i}$

47. Use De Moivre's Theorem with $n = 3$ to express $\cos 3\theta$ and $\sin 3\theta$ in terms of $\cos \theta$ and $\sin \theta$.

- 48.** Use Euler's formula to prove the following formulas for $\cos x$ and $\sin x$:

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \quad \sin x = \frac{e^{ix} - e^{-ix}}{2i}$$

- 49.** If $u(x) = f(x) + ig(x)$ is a complex-valued function of a real variable x and the real and imaginary parts $f(x)$ and $g(x)$ are differentiable functions of x , then the derivative of u is defined to be $u'(x) = f'(x) + ig'(x)$. Use this together with Equation 7 to prove that if $F(x) = e^{rx}$, then $F'(x) = re^{rx}$ when $r = a + bi$ is a complex number.

- 50.** (a) If u is a complex-valued function of a real variable, its indefinite integral $\int u(x) dx$ is an antiderivative of u . Evaluate

$$\int e^{(1+i)x} dx$$

- (b) By considering the real and imaginary parts of the integral in part (a), evaluate the real integrals

$$\int e^x \cos x dx \quad \text{and} \quad \int e^x \sin x dx$$

- (c) Compare with the method used in Example 4 in Section 6.1.

Appendix C: Linear Algebra

(Starting in the middle of the next page.)

1.1.4 Joint and Conditional Probabilities

Sometimes we're interested in the probability of more than one thing. For example, we want to know the probability that a particle has a particular value of x *and* a particular value of y (i.e., it has both of these values simultaneously). We denote this probability by $P(x, y)$, and refer to it as the joint probability of x and y .

The joint probability density $p(x, y)$, has the property that integrating with respect to one variable yields the probability density of the other. For example,

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy. \quad (1.25)$$

Here we say that $p(x)$ is the marginal probability density of x . Averages are obtained by integrating with respect to both variables:

$$\langle f(x, y) \rangle = \int_{-\infty}^{\infty} f(x, y) p(x, y) dx dy. \quad (1.26)$$

In some situations we want to know the probability of obtaining x , given a particular value of y . We denote this probability by $P(x|y)$, and call it a conditional probability, because it represents the probability of x conditioned on y . The joint and conditional probabilities are related by

$$P(x, y) = P(x|y) P(y). \quad (1.27)$$

For more details see ref. [1.1], sec. 3.7.

1.2 LINEAR ALGEBRA

1.2.1 Vectors and Basis Sets

We'll use what you already know about vectors from first-year physics to develop a vocabulary for linear algebra, and then we'll build on that.

We can express a vector \mathbf{a} (we'll denote vectors with bold type), that "lives" in the x - y plane as

$$\mathbf{a} = a_x \mathbf{u}_x + a_y \mathbf{u}_y, \quad (1.28)$$

where \mathbf{u}_x and \mathbf{u}_y are dimensionless unit vectors that point in the positive x - and y -directions respectively.³ Any vector in the x - y plane can be expressed by giving its components (coordinates) in the \mathbf{u}_x and \mathbf{u}_y directions; in eq. (1.28) the components are a_x and a_y . The vectors \mathbf{u}_x and \mathbf{u}_y are called basis vectors, and they make up a basis set. This basis set contains two vectors, because it is a 2-D vector space (an N -dimensional vector space must have N basis vectors).

³ We are not using the more familiar notion of \hat{i} and \hat{j} , or \hat{x} and \hat{y} , to denote the unit vectors because we will reserve the caret symbol $\hat{\cdot}$ for something else.

Vectors are a shorthand notation. The symbol a is a convenient way to denote an object that has multiple components. To specify the 2-D vector a we need to specify two numbers, its components a_x and a_y . Equation (1.28) is one particular notation for doing this, but there are others. For example, we could specify a as a row vector

$$a = (a_x, a_y), \quad (1.29)$$

or a column vector

$$a = \begin{pmatrix} a_x \\ a_y \end{pmatrix}. \quad (1.30)$$

As long as we know what the components of a are, we know what the vector a is. There are different notations for a , but they all represent the same vector.⁴

In terms of column vectors, the basis vectors in the x - y plane can be written

$$u_x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, u_y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (1.31)$$

The column vector equivalent of eq. (1.28) is thus

$$a = a_x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a_y \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} a_x \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ a_y \end{pmatrix} = \begin{pmatrix} a_x \\ a_y \end{pmatrix}. \quad (1.32)$$

The row vector equivalent can be written out in a similar manner.

One nice thing about the notation of eq. (1.28) is that it explicitly contains the basis vectors u_x and u_y ; there is no ambiguity about what the basis vectors are. However, the row and column vectors of eqs. (1.29) and (1.30) do not explicitly reference the basis vectors. The reader needs to know what basis is being used, because if the basis changes the row and column vectors change. This is best illustrated with a specific example.

Let's say that in the x - y coordinate system

$$a = 2u_x + 2u_y = \begin{pmatrix} 2 \\ 2 \end{pmatrix}. \quad (1.33)$$

This vector is displayed graphically in fig. 1.2(a). We can also express a in the x' - y' coordinate system, which is rotated from the x - y system by 45° , as shown in fig. 1.2(b). In this coordinate system

$$a = 2\sqrt{2}u'_x = \begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}. \quad (1.34)$$

4. In eqs. (1.29) and (1.30) we didn't use $=$, but rather the symbol \doteq . We'll use \doteq to denote "is represented by" as opposed to "is equal to". The reasons for doing this will be detailed below. We're borrowing this notation from ref. [1.2].

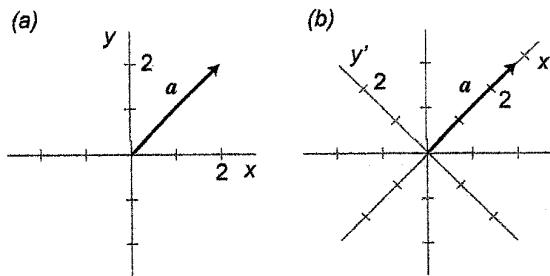


Fig 1.2 The vector a represented in: (a) the x - y coordinate system, (b) the x' - y' coordinate system.

As can be seen in fig. 1.2, a has not changed—it has the same length and points in the same direction. In eqs. (1.33) and (1.34) we have merely expressed a using two different basis sets. If we use the unit vector notation there is no ambiguity about what coordinate system we're using. If we use the column vector notation, however, there is no indication of what basis we're talking about. If we simply see $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$, how do we know whether this is expressed in the x - y basis or the x' - y' basis? From now on, if there is potential for confusion, we will attempt to remove this ambiguity by placing a subscript on the vectors to indicate which basis they are being expressed in. For example, we will write these vectors as $\begin{pmatrix} 2 \\ 2 \end{pmatrix}_{x,y}$ and $\begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}_{x',y'}$.

The difference between a vector, and its representation as a row or column vector, is why we're using the symbol \doteq to mean “is represented by”. The vector itself is always the same (i.e., it is independent of the representation), but it is represented differently depending upon which basis we use.

1.2.2 The Inner Product

We know that we can write the dot product of the vectors a and b as

$$a \cdot b = a_x b_x + a_y b_y + a_z b_z. \quad (1.35)$$

The dot product is a way of “multiplying” two vectors. It's also called the scalar product because the result of this operation is a scalar, not a vector. An operation which combines two vectors to produce a scalar is generally referred to as an inner product. To compute the inner product we multiply the components of the two vectors, and then add up all the products. In terms of row and column vectors, the inner product of a and b would be written as

$$\mathbf{a} \cdot \mathbf{b} = (a_x, a_y, a_z) \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = a_x b_x + a_y b_y + a_z b_z. \quad (1.36)$$

Since we'll be dealing with vector spaces that are more general than three-dimensional real space, it won't always be convenient to use the subscripts x , y , and z to denote the coordinates. Instead, we'll often use numbers to denote the coordinates. For example

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}. \quad (1.37)$$

We've written the inner product of eq. (1.36) in terms of the product of a row vector and a column vector, with the row vector written on the left. This is the way we have to write inner products in terms of row and column vectors. We cannot directly take the inner product of two row or two column vectors. If we have two column vectors, for example, we must first convert one to a row vector before taking the inner product. There are two steps to doing this: first, we write the elements of the column vector as a row vector; then we take the complex conjugate of each of the elements. The row vector corresponding to the column vector in eq. (1.37) is

$$\mathbf{a} = (a_1^*, a_2^*, a_3^*). \quad (1.38)$$

We convert a row vector to a column vector in the same manner. You're probably used to thinking about real vectors that live in real space, so you wouldn't ordinarily think about taking the complex conjugate. However, in quantum mechanics the vectors we'll deal with are in general complex, and we have to keep this in mind.

Why do we take the complex conjugate when switching to a row vector from a column vector, or vice versa? Imagine taking the inner product of a vector with itself. To do this we need two versions of the vector, one a row and the other a column. The inner product is then

$$(a_1^*, a_2^*, a_3^*) \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = a_1^* a_1 + a_2^* a_2 + a_3^* a_3 = |a_1|^2 + |a_2|^2 + |a_3|^2. \quad (1.39)$$

The complex conjugate ensures that we get a real number when we take the inner product of a vector with itself. The square root of the inner product of a vector with itself is called the norm of the vector. The norm is a measure of the "length" of the vector, which is why we desire that it always be a positive number. A vector is normalized if its norm (or consequently its norm squared) is equal to 1.

Ordinarily we think of the components of a vector as being the elements of its column vector representation. With this in mind, the inner product of two vectors, \mathbf{a} and \mathbf{b} , in a general N -dimensional space can be written as

$$\left(a_1^*, \dots, a_N^* \right) \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} = \sum_{i=1}^N a_i^* b_i. \quad (1.40)$$

If we already know the row vector representation, we don't need to take the complex conjugate *again* to compute the inner product. We only need to do this when switching from a column vector to a row vector.

In ordinary three-dimensional (3-D) space we think of vectors as being orthogonal if they make a 90° angle with respect to each other. Equivalently, we say that two vectors are orthogonal if their dot product is 0. In the more general vector spaces we will be talking about in quantum mechanics, two vectors are orthogonal if their inner product is equal to 0.

Again, thinking about 3-D space, we almost always work in coordinate systems where our basis vectors are mutually orthogonal, and we also find it convenient to use normalized basis vectors. If we have a basis set in which all of the vectors are mutually orthogonal, and normalized, we have an orthonormal basis set. For example, the unit vectors \mathbf{u}_x , \mathbf{u}_y , and \mathbf{u}_z in 3-D are an orthonormal basis set.

1.2.3 Matrices

In our discussion of matrices we'll focus on square matrices (i.e., matrices which have the same number of rows and columns), because 99% of the time in quantum mechanics that's what we deal with. I'll assume that you know how to take the determinant of such a matrix. I'll also assume that you know how to multiply a matrix and a vector, or two matrices. However, to be clear on notation we'll present some expressions, which you might not be familiar with, that describe these processes.

We'll use M_{ij} to denote the element in the i^{th} row and j^{th} column of the matrix $\bar{\mathbf{M}}$ (we'll use bold with an overbar to denote a matrix). Recall that the product of a matrix and column vector yields another column vector. In N -dimensions, the components of \mathbf{b} , which comes from the operation $\mathbf{b} = \bar{\mathbf{M}}\mathbf{a}$, can be written as the inner product of the rows of $\bar{\mathbf{M}}$ with \mathbf{a} :

$$b_i = \sum_{j=1}^N M_{ij} a_j. \quad (1.41)$$

Since knowing the components determines the vector, eq. (1.41) expresses the multiplication of a matrix and a column vector. Similarly, we can express the elements of the matrix product $\bar{\mathbf{M}} = \bar{\mathbf{A}}\bar{\mathbf{B}}$ in terms of the inner product of the i^{th} row of the first matrix with the j^{th} column of the second:

$$M_{ij} = \sum_{k=1}^N A_{ik} B_{kj}. \quad (1.42)$$

You're probably already aware, and in the problems you'll confirm, that in general $\bar{A}\bar{B} \neq \bar{B}\bar{A}$. The order in which we multiply matrices *matters*—matrix multiplication is not commutative. More typically, we'll say that matrices don't always commute. This has important implications in quantum mechanics, as we'll discuss later.

1.2.4 Eigenvalues and Eigenvectors

The following is a mathematical problem that arises frequently, often as the result of a physical problem: We are given a matrix \bar{M} , and we want to find the vectors x and constants λ that are solutions of the equation

$$\bar{M}x = \lambda x. \quad (1.43)$$

In other words, there are some special vectors such that when we multiply them by the matrix \bar{M} , give us back a scaled version of themselves. The vectors that are solutions to eq. (1.43) are called the eigenvectors of \bar{M} , and the corresponding constants are called the eigenvalues of \bar{M} .⁵

How do we solve this problem? Start by rewriting eq. (1.43) as

$$\bar{M}x - \lambda x = 0. \quad (1.44)$$

In order to factor out the vector x , we need to introduce the identity matrix \bar{I} , which has 1s on the diagonal, and 0s everywhere else. In 3-D, for example, the identity is given by

$$\bar{I} \doteq \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (1.45)$$

Another way to express the identity matrix is to say that its elements are equal to the Kronecker delta δ_{ij} , which is defined as

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (1.46)$$

The identity matrix is useful because multiplying it times any vector (or matrix) simply returns the same vector (or matrix). For example, in 2-D

$$\bar{I}x \doteq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \doteq x. \quad (1.47)$$

Using this fact, we can insert the identity into eq. (1.44) and obtain

5. More generally this problem is expressed in terms of linear operators, as opposed to matrices, as we'll discuss in later chapters.

$$\bar{\mathbf{M}}\mathbf{x} - \lambda\mathbf{x} = \bar{\mathbf{M}}\mathbf{x} - \lambda\bar{\mathbf{1}}\mathbf{x} = (\bar{\mathbf{M}} - \lambda\bar{\mathbf{1}})\mathbf{x} = 0. \quad (1.48)$$

Instead of continuing on generally, we'll specialize to 2-D, so we can explicitly write eq. (1.48) as

$$\begin{pmatrix} M_{11} - \lambda & M_{12} \\ M_{21} & M_{22} - \lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (1.49)$$

This is equivalent to a system of 2 equations

$$\begin{aligned} (M_{11} - \lambda)x_1 + M_{12}x_2 &= 0, \\ M_{21}x_1 + (M_{22} - \lambda)x_2 &= 0, \end{aligned} \quad (1.50)$$

in 3 unknowns: λ , x_1 and x_2 . There is a solution to this problem if and only if the determinant of the matrix in eq. (1.49) is 0:

$$\begin{vmatrix} M_{11} - \lambda & M_{12} \\ M_{21} & M_{22} - \lambda \end{vmatrix} = (M_{11} - \lambda)(M_{22} - \lambda) - M_{12}M_{21} = 0. \quad (1.51)$$

We'll just outline the rest of the procedure before doing a specific example. Begin by noticing that eq. (1.51) is 2nd-order in λ . This means that there are 2 solutions, which we'll call λ_a and λ_b . Once these eigenvalues have been found, each will have its own corresponding eigenvector, \mathbf{x}_a or \mathbf{x}_b . To find the eigenvectors, substitute an eigenvalue into the original equation, then solve for the corresponding eigenvector. For example:

$$\begin{pmatrix} M_{11} - \lambda_a & M_{12} \\ M_{21} & M_{22} - \lambda_a \end{pmatrix} \begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} = 0. \quad (1.52)$$

This is equivalent to two linear equations in two unknowns, x_{a1} and x_{a2} , so we can solve for them

$$\begin{aligned} (M_{11} - \lambda_a)x_{a1} + M_{12}x_{a2} &= 0, \\ M_{21}x_{a1} + (M_{22} - \lambda_a)x_{a2} &= 0. \end{aligned} \quad (1.53)$$

In a general N -dimensional problem there will be N eigenvalues and N eigenvectors.

There's one last trick. The equations that determine the eigenvectors [e.g., eq. (1.53)] are *not* independent (the solutions for the eigenvalues ensure that this is the case). This means that there is no unique solution to the problem; we can solve for x_{a2} in terms of x_{a1} , or vice versa, but neither is uniquely determined. That this must be so is seen in the fact that if \mathbf{x}_a is a solution to our original problem, eq. (1.43), then any constant times \mathbf{x}_a is also a solution—there are an infinite number of \mathbf{x}_a 's corresponding to λ_a . We are thus free to impose one more constraint in order to obtain a unique solution. Typically we choose to normalize the eigenvectors. This means

$$\begin{pmatrix} x_{a1}^* & x_{a2}^* \end{pmatrix} \begin{pmatrix} x_{a1} \\ x_{a2} \end{pmatrix} = |x_{a1}|^2 + |x_{a2}|^2 = 1. \quad (1.54)$$

It's probably easiest to learn how to find eigenvalues and eigenvectors using an example.

EXAMPLE 1.4

Find the eigenvalues and eigenvectors of

$$\bar{\mathbf{M}} = \begin{pmatrix} 0 & 1 \\ -2 & 3 \end{pmatrix}. \quad (1.55)$$

First we find the eigenvalues by subtracting λ from each of the diagonal terms, and setting the determinant equal to 0 [eq. (1.51)]:

$$\begin{vmatrix} 0-\lambda & 1 \\ -2 & 3-\lambda \end{vmatrix} = 0, \quad (1.56)$$

which yields

$$-\lambda(3-\lambda) + 2 = \lambda^2 - 3\lambda + 2 = 0. \quad (1.57)$$

Solving this quadratic equation yields two eigenvalues:

$$\lambda_{\pm} = \frac{-(-3) \pm \sqrt{(-3)^2 - 4(1)(2)}}{2(1)}, \quad (1.58)$$

$$\lambda_+ = 2, \quad \lambda_- = 1. \quad (1.59)$$

To find the eigenvector corresponding to $\lambda_+ = 2$, we know that [eq. (1.52)]

$$\begin{pmatrix} 0-2 & 1 \\ -2 & 3-2 \end{pmatrix} \begin{pmatrix} x_{+1} \\ x_{+2} \end{pmatrix} = 0. \quad (1.60)$$

This yields 2 equations in 2 unknowns:

$$\begin{aligned} -2x_{+1} + x_{+2} &= 0, \\ -2x_{+1} + x_{+2} &= 0. \end{aligned} \quad (1.61)$$

Notice that these are the same equation, so we solve it to obtain $x_{+2} = 2x_{+1}$. An eigenvector corresponding to $\lambda_+ = 2$ is thus:

$$\mathbf{x}_+ \doteq \begin{pmatrix} x_{+1} \\ 2x_{+1} \end{pmatrix} = x_{+1} \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \quad (1.62)$$

We can normalize this using

$$(x_{+1})^2 (1^2 + 2^2) = 1, x_{+1} = \frac{1}{\sqrt{5}}. \quad (1.63)$$

So, one eigenvalue-eigenvector pair is

$$\lambda_+ = 2, \mathbf{x}_+ \doteq \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \quad (1.64)$$

The eigenvector corresponding to $\lambda_- = 1$ is found in a similar fashion; it is

$$\mathbf{x}_- = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (1.65)$$

1.3 References

- [1.1] I. Miller and M. Miller, *John E. Freund's Mathematical Statistics with Applications*, 7th ed. (Pearson Prentice Hall, Upper Saddle River, NJ, 2004).
- [1.2] J.J. Sakurai and J. Napolitano, *Modern Quantum Mechanics*, 2nd ed. (Addison-Wesley, Boston, MA, 2011).

1.4 PROBLEMS

Note: In several of these problems you will be asked to calculate the variance, or the standard deviation. If you use your calculator or a spreadsheet, make sure you know whether it uses $1/N$ or $1/(N-1)$. If it uses $1/(N-1)$, convert to $1/N$.

Data A (x_i): 10, 13, 14, 14, 6, 8, 7, 9, 12, 14, 13, 11, 10, 7, 7

Data B (x_i, y_i): (3,4), (5,8), (4,4), (8,5), (3,5), (4,5), (5,8), (8,5), (8,4), (3,4), (3,8), (4,8)

- 1.1 Calculate the mean and variance of Data A directly from the data.
- 1.2 Create a histogram of Data A, and from it estimate the probability distribution. Use min and max values for the data range of 5 and 14, and use 5 bins (5 and 6 go in the first bin; label it $x_i=5.5$, etc.).
- 1.3 Use the probability distribution from problem 1.2 to calculate the mean of