

1

Quantum measurement theory

1.1 Introduction and overview

Much that is studied in quantum measurement theory, and thus in this book, is very different from that studied in its classical counterpart, Bayesian inference. While quantum measurement theory is in one sense a minimal extension of the latter to quantum states, quantum measurements cause dynamical changes that never need appear in a classical theory. This has a multitude of ramifications, many of which you will find in this book. One such consequence is a limit to the information that can be obtained about the state of a quantum system, and this results in relationships between information gained and dynamics induced that do not exist in classical systems. Because of the limits to information extraction, choosing the right kinds of measurements becomes important in a way that it never was for classical systems. Finding the best measurements for a given purpose, and working out how to realize these measurements, is often nontrivial.

The relationship between information and disturbance impacts the control of quantum systems using measurements and feedback, and we discuss this application in Chapter 5. The ability to extract information about quantum systems is also important in using quantum systems to make precise measurements of classical quantities such as time, acceleration, and the strengths of electric and magnetic fields. This subject is referred to as quantum metrology, and we discuss it in Chapter 6.

The dynamics generated by quantum measurements is as rich as the usual unitary evolution, and quite distinct from it. While the latter is linear, the former is both nonlinear and random (stochastic). The nonlinear dynamics due to quantum measurement, as we will see in Chapter 4, exhibits all the chaotic dynamics of nonlinear classical systems. This is made more interesting by the fact that all processes involving measurements, as we will see in Chapter 1, can be rewritten (almost) entirely as unitary and thus linear processes.

The role that measurements can play in thermodynamics is also an interesting one, especially because of the close connection between the entropy of statistical mechanics and the entropy of information theory. While this is not a special feature of quantum, as opposed to classical measurements, we spend some time on it in Chapter 4 both because of its fundamental nature, and because it is an application of measurement theory to the manipulation of heat and work.

There is another important connection between quantum measurement theory and thermodynamics: thermal baths that induce thermalization and damping also carry away a continuous stream of information from the system with which they interact. Because of this, thermal baths can mediate continuous measurements, and there is a useful overlap between the descriptions of the two. Further, the continuous measurements induced by a system's environment are sufficient to induce the quantum-to-classical transition, in which the trajectories of classical dynamics emerge as a result of quantum mechanics.

It is now possible to construct and manipulate individual quantum systems in the laboratory in an increasingly wide range of physical settings. Because of this, we complete our coverage of measurement theory by applying it to the measurement and control of a variety of concrete mesoscopic systems. We discuss nano-electromechanical systems, in which superconducting circuits are coupled to nano-mechanical resonators, and opto-mechanical systems in which superconducting circuit elements are replaced by the modes of optical resonators. In Chapter 7 we introduce these systems and show how to determine their Hamiltonians and the interactions between them. In this chapter we also explain how the continuous quantum measurements introduced in Chapter 3 are realized in these systems, and rephrase them in the language usually used by experimentalists, that of amplifiers. In Chapter 8 we consider a number of examples in which the above systems are controlled, including realizations of the feedback control techniques described in Chapter 5.

Despite its importance, quantum measurement theory has had an uneasy relationship with physicists since the inception of quantum mechanics. For a long time the tendency was to ignore its predictions as far as measurements on individual systems were concerned, and consider only the predictions involved with averages of ensembles of systems. But as experimental technology increased, and experimentalists were increasingly able not only to prepare and measure individual quantum systems in well-defined states with few quanta but to make repeated measurements on the same system, physicists were increasingly forced to use the full predictions of measurement theory to explain observations.

The uneasiness associated with measurement theory stems from a philosophical issue, and this has disturbed physicists enough that many still seem uncomfortable with using the language of measurement in quantum mechanics, and assigning the knowledge gained to an observer. In fact, as will be explained in due course, all processes that involve quantum measurements can be described without them, with the possible exception of a single measurement relegated to the end of the process. In using this second description the language changes, but the two descriptions are completely equivalent. Thus the use of the language of measurement theory is entirely justified. What is more, analysis that makes explicit use of measurement theory is often vastly more convenient than its unitary equivalent, providing a strong practical reason for using it. This should not be taken to mean that the philosophical problem has gone away – far from it. In fact, while the focus of this book is applications, we feel that no education in quantum measurement theory is quite complete without an understanding of “the measurement problem,” and because of this we include a discussion of it at the end of Chapter 4.

We do not cover all the applications of quantum measurement theory, most notably those to do with quantum information theory. The latter is now itself a large subfield of physics, and there are a number of textbooks devoted to it. While our goal is to focus on applications that are not usually covered in books on quantum information, given the fundamental connection between measurement and information the two subjects are intertwined. A number of concepts from information theory are therefore invaluable in broader applications of measurement, and so we discuss information theory in Chapter 2. We do not cover “foundational” questions in quantum theory, with the sole exception of the so-called “quantum measurement problem.” Another topic within measurement theory that we do not discuss is that termed “weak values,” not to be confused with weak measurements which we treat in detail.

Quantum measurement theory is a weird and wonderful subject. I hope I have whetted your appetite, and that you find some topics in this book as stimulating and fascinating as I have.

A guide to this book

Different readers will want to take different pathways through the text. Many experimentalists, for example, may wish to learn how to describe weak or continuous measurements in their experiments without wading through Chapters 1 and 3. Such readers can start with Appendix B, and possibly Appendix C, from which a number of parts of the book are accessible. Experimentalists working with mesoscopic systems can then proceed to Section 7.7.1 that describes measurements using the language of amplifiers, and Section 3.1.4 that shows how to calculate the spectrum of a measurement of a linear quantum system using an equivalent classical model. In fact, Section 7.7.1 can be read first if desired, as it does not require the theoretical language used for continuous measurements. Section 5.4.3 is also accessible from the above appendices, and deals with the feedback control of a single qubit. Experimentalists working with quantum and atom optics could instead proceed to Sections 3.3 and 3.4 that deal with optical cavities, photon counting, and optical homodyne detection, although these sections do also require Chapter 1 up to and including Section 1.3.3.

The first parts of Chapter 7 and much of Chapter 8, introducing mesoscopic circuits and optomechanical systems, do not require any previous material. Sections 7.7.3 and 8.2 are the exceptions, requiring either Appendix B or Chapters 1 and 3.

Students who want to learn the broad fundamentals of measurement theory should start with Chapters 1 and 2, and proceed to Chapter 3 if they are interested in the details of continuous measurements. Section 4.3 of Chapter 4 is also core reading on the relationship between continuous measurements and open systems.

Having studied Chapters 1, 2, and 3 all of the remainder of the text is accessible and topics can be selected on preference. We now list the latter sections that do not require Chapter 3, since these can be included in a course of study that does not involve continuous measurements: Sections 4.1.2 and 4.1.3 on Landauer’s erasure principle and Maxwell’s demon; parts of Section 5.3 on feedback control; all but Section 6.2.2 in Chapter 6 on

metrology; Sections 7.1 through 7.6, and Sections 8.1 and 8.3 on controlling mesoscopic systems.

Some assumptions and terminology

This text is for graduate physics students, so we assume that the reader is familiar with quantum mechanics, the basics of probability theory, and various mathematical concepts such as Fourier transforms and δ -functions. Everything that the reader needs to know about probability theory and Fourier transforms can be found in Chapter 1 of reference [288] or Chapter 4 of reference [592] and Chapter 1 of reference [591]. We also recommend Jaynes' landmark book on probability theory and its role in reasoning and science [311].

In mathematics texts it is usual to denote a random variable as a capital letter, say X , and the variable denoting one of the values it can take as the corresponding lower case letter, x . This provides technical precision, since the concept of a random variable, and the concept of one of the values it can take, are distinct. However, physicists tend to use the same symbol to denote both things, because it causes no confusion in practice. This is the style I prefer, so I use it here.

We physicists often use the term “probability distribution” as synonymous with “probability density,” whereas mathematicians use the former term to mean the anti-derivative of the latter. Defining “probability distribution” to mean “probability density” is useful, because standard English usage prevents us from using “probability density” for discrete distributions. For this reason we will use the term *probability distribution* to mean probability density.

To refer to a set of quantities, for example a set of probabilities p_j indexed by j , in which j takes integer values in some range, we will use the shorthand $\{p_j\}$. The locations of the definitions of all acronyms we use can be found in the index.

1.2 Classical measurement theory

Before learning quantum measurement theory, it is valuable to understand classical measurement theory, as the two are very closely connected. Classical measurement theory, also known as Bayesian statistical inference, tells us how our knowledge about the value of some quantity, x , changes when we obtain a piece of data relating to x . To understand how this works, we need to know first how to describe the knowledge we have regarding x . We will assume that x is some variable that can take any real value, but more generally it could be a vector of discrete or continuous variables. Since we need to make a measurement to determine x , we must be uncertain about its value. Our knowledge about x is therefore captured by a probability distribution, $P(x)$, for the values of x . This probability distribution tells us, based on the information currently available, the likelihood that x will have various values, and overall how certain, or uncertain, we are about x . This distribution is called our *state-of-knowledge* of x .

To determine how our state-of-knowledge changes when we obtain a piece of data y , we have to know how y is related to x . To be able to describe measurements, this relationship

must be probabilistic: if y were deterministically related to x (meaning that y was a function of x), then we could determine x precisely once we knew y , simply by inverting the function. This is not what happens when we make measurements; measurements are never perfectly precise. After we have made a measurement of x , giving a result y , we are always left with some uncertainty about x .

Consider measuring the length of an object with a ruler. In this case the result of the measurement is equal to the true length plus a random error. Thus, given the true length of the object (the value of x) there is a probability distribution for the result y . The probability distribution for y is peaked at $y = x$. Because this probability distribution for y depends on (is conditional upon) the value of x , it is called a *conditional* probability distribution for y , and is written $P(y|x)$ (this is read as “ P of y given x ”). The conditional probability for the measurement result (the data) y , given the value of the quantity to be measured, x , completely defines the measurement. This conditional probability is determined by the physical procedure used to make the measurement, and is referred to as the **likelihood function** for the measurement. So if one wishes to obtain the likelihood function for a given measurement process, one must first work out how the value of the thing to be measured leads to the measurement result. If y is the measurement result, and x is the quantity to be measured, then the process by which y leads to x can often be written in the form $y = f(x) + \text{a random variable}$, where f is a deterministic function. The likelihood function can then be determined directly from this relationship.

To determine how our state-of-knowledge regarding x , $P(x)$, changes when we obtain the value of y , we use the relationships between the joint probability for two random variables x and y , $P(x, y)$, and the conditional probabilities $P(y|x)$ and $P(x|y)$. These relationships are

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x). \quad (1.1)$$

Here $P(y)$ is the probability distribution for y irrespective of the value of x (also called the *marginal* distribution for y), and is given by

$$P(y) = \int_{-\infty}^{\infty} P(x, y) dx. \quad (1.2)$$

Now $P(x)$ is our state-of-knowledge of x prior to making the measurement, and it is therefore the probability density for x irrespective of the value of y . It is therefore also the marginal probability for x , and thus given by

$$P(x) = \int_{-\infty}^{\infty} P(x, y) dy. \quad (1.3)$$

While the relationships in Eq. (1.1) are fairly intuitive, they are explained further in, for example, references [288] and [592].

Rearranging Eq. (1.1) we obtain the famous relationship known as *Bayes' theorem*, being

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}. \quad (1.4)$$

Upon examining Eq. (1.4) we will see that it tells us exactly how to change our state-of-knowledge when we obtain the measurement result y . First note that since $P(x|y)$ must be normalized (that is, its integral over x must be unity), the value of $P(y)$ on the bottom line is completely determined by this normalization. We can therefore write Bayes' theorem as

$$P(x|y) = \frac{P(y|x)P(x)}{\mathcal{N}}, \quad \text{where } \mathcal{N} = \int_{-\infty}^{\infty} P(y|x)P(x)dx = P(y). \quad (1.5)$$

We see that on the right-hand side (RHS) we have our state-of-knowledge of x before we obtain the data y , and on the left-hand side (LHS) the probability for x *given* that value of y . The LHS is therefore our state-of-knowledge after obtaining the value of y . In Eq. (1.5) $P(x)$ is called the **prior** probability, and $P(x|y)$ the **posterior** probability. So Bayes' theorem tells us that to obtain our new state-of-knowledge once we have made our measurement: we simply multiply our current state-of-knowledge by the likelihood function $P(y|x)$, and normalize the result. Note that the prior is simply the marginal (overall) distribution of x . The relationship given by Eq. (1.5), Bayes' theorem, is the fundamental theorem of classical measurement theory.

1.2.1 Understanding Bayes' theorem

While Bayes' theorem (Eq. 1.5) is simple to derive, to obtain a direct understanding of it requires a bit more work. To this end consider a measurement of a discrete variable, x , in which x has only two values. We will call these values **0** and **1**. Our measurement will also have only two outcomes, which we will denote by $y = 0$ and $y = 1$. In our discussion of Bayes' theorem above, we assumed that x was continuous, so let us take a minute to re-orient our thinking to discrete variables. In the present case our state-of-knowledge (our prior), $P(x)$, has only two values, being $P(\mathbf{0})$ and $P(\mathbf{1})$ (and, of course, $P(\mathbf{0}) + P(\mathbf{1}) = 1$). The conditional probability, $P(y|x)$, also has only two values for each value of x . If we make the measurement and obtain the result $y = 1$ (for example), then Bayes' theorem tells us that our posterior is given by

$$P(x|1) = \frac{P(1|x)P(x)}{\sum_{x'} P(1|x')P(x')} = \frac{P(1|x)P(x)}{P(1|\mathbf{0})P(\mathbf{0}) + P(1|\mathbf{1})P(\mathbf{1})}. \quad (1.6)$$

We now wish to obtain a better understanding of this expression.

To do so let us choose a specific likelihood function for the measurement. This likelihood function is given in Table 1.1, and contains the parameter α . If α is close to unity, then the two values of x give very different distributions for the measurement result y , and in this

Table 1.1. *The likelihood function for a simple “two-outcome” measurement. The body of the table gives the probability for y given x .*

$P(y x)$	$x = 0$	$x = 1$
$y = 0$	α	$1 - \alpha$
$y = 1$	$1 - \alpha$	α

case we would expect the measurement to tell us a lot about x . Conversely, if α is close to $1/2$, then the reverse is true. For the sake of concreteness let us choose $0.5 < \alpha < 1$. With this choice, if $x = 0$ then it is more likely that we will get the result $y = 0$, and if $x = 1$ it is more likely that we will get $y = 1$.

Now assume that we initially know nothing about x , so that our prior state of knowledge is $P(0) = 0.5 = P(1)$. What happens when we make the measurement and get the result $y = 1$? Since our prior is flat, by which we mean that it does not change with x , it cancels on the top and bottom lines of Bayes’ theorem, telling us that our posterior is simply the likelihood function, normalized if necessary:

$$P(x|y) = \frac{P(y|x)}{\sum_{x'} P(y|x')} = \frac{P(y|x)}{P(y|0) + P(y|1)}. \tag{1.7}$$

Our posterior is thus $P(0|1) = 1 - \alpha$ and $P(1|1) = \alpha$. So it is now more likely that $x = 1$ than $x = 0$. This is indeed intuitively reasonable. The likelihood function tells us that if x were to be 0 then it would be less likely that we would get the result 1 , so it is reasonable that since we obtained the result 1 , it is more likely that $x = 1$ than that $x = 0$.

The above discussion shows that if the prior is uniform, Bayes’ theorem tells us that *the values of x that are more likely are those for which the result we have obtained is the more likely outcome*.

Now let us examine why we need to include the prior, in addition to the likelihood function, when calculating our posterior. This is clearest when the prior is strongly weighted toward one value of x . Consider a situation in which the prior is $P(0) = 0.999$ and $P(1) = 0.001$. This means that in the absence of any further data, on average x will only be equal to 1 one time in a thousand cases. We now consider a slightly different two-outcome measurement from the one above, as this will make the logic simple. The likelihood function for the new measurement is given in Table 1.2, and in words is as follows: if $x = 1$ then y is always equal to 1 . If $x = 0$, then $y = 0$ with probability 0.999 and $y = 1$ only one time in a thousand. This means that, if the prior is flat, upon obtaining the result $y = 1$ the value of x would be equal to 1 approximately nine hundred and ninety-nine times out of a

Table 1.2. *The likelihood function for our second “two-outcome” measurement. The body of the table gives the probability for y given x .*

$P(y x)$	$x = 0$	$x = 1$
$y = 0$	0.999	0
$y = 1$	0.001	1

thousand:

$$P(x = 1|y = 1) = \frac{P(1|1)}{P(1|0) + P(1|1)} = \frac{0.999}{1.001}. \tag{1.8}$$

So what is the case when the prior is highly weighted toward $x = 0$, as described above? Well, in this case x is equal to 1 only one time in a thousand. Now, if we get the result $y = 1$ there are two possibilities. Either $x = 1$, which happens one time in one thousand, or $x = 0$ and we got the result $y = 1$ anyway, which also happens approximately one time in a thousand. (The precise figure is the frequency of $x = 0$ multiplied by the frequency of the result $y = 1$ given $x = 0$, which is $0.999 \times 0.001 \approx 1/1001$.) Thus the result $y = 1$ happens approximately one time in 500, and half of these are due to $x = 0$, and half due to $x = 1$. So when we obtain the result $y = 1$, there is only a 50% chance that $x = 1$. This is, of course, exactly what Bayes’ theorem tells us; by multiplying the likelihood function that weights $x = 1$ very highly, by the prior that weights $x = 0$ very highly, we obtain an approximately flat posterior.

The example we have just considered applies directly to a real and very important situation: testing for the HIV virus. Each test is pretty reliable, giving a false positive only about one time in a thousand. On that basis alone one might think that when a result comes back positive, there is little reason to perform a follow-up test to confirm it. But this is very wrong. Since very few patients have HIV, false positives come up just as frequently as real positives. Thus, whenever a positive test result comes back, it is essential to do a follow-up test to check that it is not a false positive. Bayesian inference, and thus measurement theory, is therefore crucial in real-world problems.

To complete our discussion of Bayes’ theorem it is worth noting that our state-of-knowledge does not necessarily become more certain when we make a measurement. To take the example of the HIV test above, before we obtain the result of the test we are almost certain that the patient is HIV negative, since the vast majority of patients are. However, upon obtaining a positive test result, there is an approximately fifty-fifty chance that the patient is HIV positive. Thus, after obtaining the measurement result, we are less certain of the HIV status of the patient. Even in view of this, all classical measurements do have the property that, upon making the measurement, we become more certain *on average* of

the value of the measured quantity (where the average is taken over all the possible measurement results). We will be able to make this statement precise once we have described how to quantify the concept of information in Chapter 2.

1.2.2 Multiple measurements and Gaussian distributions

Multiple measurements

Having made a measurement of x , what happens when we make a second measurement? We might expect that we simply repeat the process of multiplying our current state-of-knowledge, now given by $P(x|y)$, by the likelihood function for the new measurement. This is correct so long as the results of the two measurements are independent, and is simple to show as follows. Let us say that we make N measurements, and the results (the data) obtained from these measurements are y_i for $i = 1, \dots, N$. Bayes' theorem tells us that

$$P(x|y_1, \dots, y_N) = \frac{P(y_1, \dots, y_N|x)P(x)}{\mathcal{N}}, \quad (1.9)$$

with $\mathcal{N} = \int_{-\infty}^{\infty} P(y_1, \dots, y_N|x)P(x)dx$. The fact that all the measurement results are independent means that

$$P(y_1, \dots, y_N|x) = P(y_1|x)P(y_2|x) \cdots P(y_N|x), \quad (1.10)$$

and with this Bayes' theorem becomes

$$\begin{aligned} P(x|y_1, \dots, y_N) &= \frac{P(y_1, \dots, y_N|x)P(x)}{\mathcal{N}} = \frac{P(y_N|x) \cdots P(y_1|x)P(x)}{\mathcal{N}} \\ &= \frac{P(y_N|x)}{\mathcal{N}_N} \cdots \frac{P(y_1|x)P(x)}{\mathcal{N}_1}, \end{aligned} \quad (1.11)$$

where $\mathcal{N}_n = P(y_n)$. So we see that each time we make another independent measurement we update our state-of-knowledge by multiplying it by the likelihood function and normalizing the result.

Pooling independent knowledge

It turns out that classical measurement theory provides us with a simple way to pool the knowledge of two observers so long as their information has been obtained independently. If two observers, A and B, have the respective states-of-knowledge $P_1(x)$ and $P_2(x)$ about a quantity x , then we can write each of these as

$$P_A(x) = \frac{P(\mathbf{y}_A|x)P_{\text{prior}}(x)}{\mathcal{N}_A}, \quad (1.12)$$

$$P_B(x) = \frac{P(\mathbf{y}_B|x)P_{\text{prior}}(x)}{\mathcal{N}_B}, \quad (1.13)$$

where $\mathbf{y}_A$ and $\mathbf{y}_B$ are the vectors of data obtained by the respective observers. Since we intend the data the observers have obtained to represent *all* the information they each have about x , $P_{\text{prior}}(x)$ is the prior that describes having no initial knowledge about x . The problem of determining such a prior can be surprisingly difficult, and we discuss it further below. If we assume for now that we know what P_{prior} is, and we choose the measure for integration over x so that $P_{\text{prior}}(x) = 1$ (that is, we absorb P_{prior} into the measure), then an observer who has access to the data of both A and B has the state-of-knowledge

$$P(x) = \frac{P(\mathbf{y}_A|x)P(\mathbf{y}_B|x)}{\mathcal{N}'} = \frac{P_A(x)P_B(x)}{\mathcal{N}}. \quad (1.14)$$

So all we have to do to pool the knowledge of two (or more) observers is to multiply their states-of-knowledge together, and normalize the result.

The ubiquity of Gaussian measurements

Now consider applying classical measurement theory to the simple example discussed above, that of measuring the length of an object with a ruler. To describe this measurement we need to decide how the result that we get, y , depends upon the true length x . We can think of y as being equal to the true length plus a random “error.” The error in a measurement is often described well by a Gaussian distribution. The reason for this is that the error is usually the result of random contributions from many different sources. When we add many independent random variables together, the central limit theorem tells us that the resulting random variable has an approximately Gaussian distribution. If the error in our measurement of x is a Gaussian with mean zero and variance V , then the probability distribution for y , given x , is a Gaussian centered at x :

$$P(y|x) = \frac{1}{\sqrt{2\pi V}} e^{-(y-x)^2/(2V)}. \quad (1.15)$$

This is the likelihood function for the measurement. If we have absolutely no knowledge of x before the measurement (never the case in reality, of course), then we can set $P(x) = 1$. Our knowledge of x after making the measurement is then simply the likelihood function, normalized so that it is a valid probability distribution over x for each value of y . In this case the normalization is already correct, and so

$$P(x|y) = \frac{1}{\sqrt{2\pi V}} e^{-(x-y)^2/(2V)}. \quad (1.16)$$

This tells us that the value of x is a Gaussian centered at y , with variance V , and thus that the most likely value of x is y , and the expectation value of x is also equal to y . It is customary to quote the error on a measured quantity as twice the standard deviation. Thus we write the value of x as $y \pm 2\sqrt{V}$.