

# 1

## A review of probability theory

In this book we will study dynamical systems driven by noise. Noise is something that changes randomly with time, and quantities that do this are called *stochastic processes*. When a dynamical system is driven by a stochastic process, its motion too has a random component, and the variables that describe it are therefore also stochastic processes. To describe noisy systems requires combining differential equations with probability theory. We begin, therefore, by reviewing what we will need to know about probability.

### 1.1 Random variables and mutually exclusive events

Probability theory is used to describe a situation in which we do not know the precise value of a variable, but may have an idea of the relative likelihood that it will have one of a number of possible values. Let us call the unknown quantity  $X$ . This quantity is referred to as a *random variable*. If  $X$  is the value that we will get when we roll a six-sided die, then the possible values of  $X$  are  $1, 2, \dots, 6$ . We describe the likelihood that  $X$  will have one of these values, say 3, by a number between 0 and 1, called the *probability*. If the probability that  $X = 3$  is unity, then this means we will *always* get 3 when we roll the die. If this probability is zero, then we will never get the value 3. If the probability is  $2/3$  that the die comes up 3, then it means that we expect to get the number 3 about two thirds of the time, if we roll the die many times.

The various values of  $X$ , and of any random variable, are an example of *mutually exclusive* events. That is, whenever we throw the die,  $X$  can have only one of the values between 1 and 6, no more and no less. Rather obviously, if the probability for  $X$  to be 3 is  $1/8$ , and for  $X$  to be 6 is  $2/8$ , then the probability for  $X$  to be *either* 3 or 6 is  $1/8 + 2/8 = 3/8$ . That is, the total probability that one of two or more mutually exclusive events occurs is the *sum* of the probabilities for each event. One usually states this by saying that “mutually exclusive probabilities sum”. Thus, if

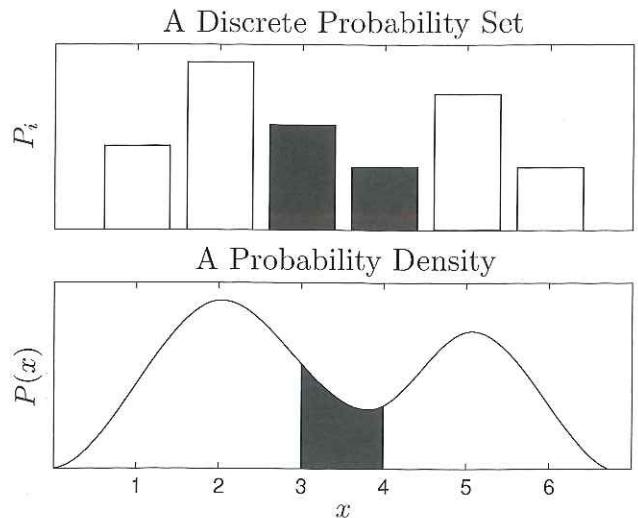


Figure 1.1. An illustration of summing the probabilities of mutually exclusive events, both for discrete and continuous random variables.

we want to know the probability for  $X$  to be in the range from 3 to 4, we sum all the probabilities for the values from 3 to 4. This is illustrated in Figure 1.1. Since  $X$  always takes a value between 1 and 6, the probability for it to take a value in this range must be unity. Thus, the sum of the probabilities for all the mutually exclusive possible values must always be unity. If the die is fair, then all the possible values are equally likely, and each is therefore equal to 1/6.

*Note:* in mathematics texts it is customary to denote the unknown quantity using a capital letter, say  $X$ , and a variable that specifies one of the possible values that  $X$  may have as the equivalent lower-case letter,  $x$ . We will use this convention in this chapter, but in the following chapters we will use a lower-case letter for both the unknown quantity and the values it can take, since it causes no confusion.

In the above example,  $X$  is a *discrete random variable*, since it takes the discrete set of values 1, ..., 6. If instead the value of  $X$  can be any real number, then we say that  $X$  is a *continuous random variable*. Once again we assign a number to each of these values to describe their relative likelihoods. This number is now a function of  $x$  (where  $x$  ranges over the values that  $X$  can take), called the *probability density*, and is usually denoted by  $P_X(x)$  (or just  $P(x)$ ). The probability for  $X$  to be in the range from  $x = a$  to  $x = b$  is now the area under  $P(x)$  from  $x = a$  to  $x = b$ . That is

$$\text{Prob}(a < X < b) = \int_a^b P(x) dx. \quad (1.1)$$

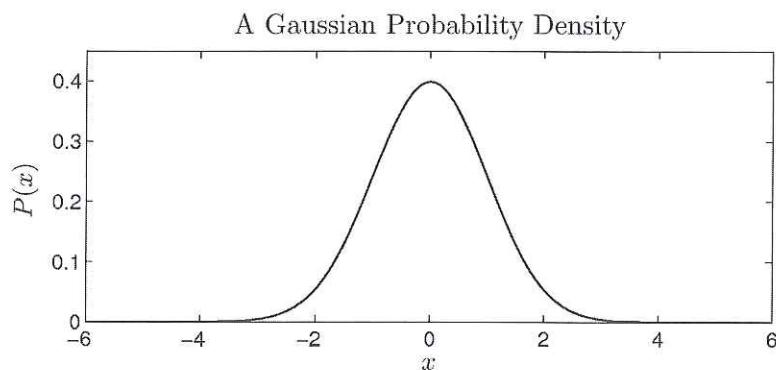


Figure 1.2. A Gaussian probability density with variance  $V = 1$ , and mean  $\langle X \rangle = 0$ .

This is illustrated in Figure 1.1. Thus the integral of  $P(x)$  over the whole real line (from  $-\infty$  to  $\infty$ ) must be 1, since  $X$  must take one of these values:

$$\int_{-\infty}^{\infty} P(x) dx = 1. \quad (1.2)$$

The average of  $X$ , also known as the *mean*, or *expectation value*, of  $X$  is defined by

$$\langle X \rangle \equiv \int_{-\infty}^{\infty} P(x) x dx. \quad (1.3)$$

If  $P(x)$  is symmetric about  $x = 0$ , then it is not difficult to see that the mean of  $X$  is zero, which is also the center of the density. If the density is symmetric about any other point, say  $x = a$ , then the mean is also  $a$ . This is clear if one considers a density that is symmetric about  $x = 0$ , and then shifts it along the  $x$ -axis so that it is symmetric about  $x = a$ : shifting the density shifts the mean by the same amount.

The *variance* of  $X$  is defined as

$$V_X \equiv \int_{-\infty}^{\infty} P(x)(x - \langle X \rangle)^2 dx = \int_{-\infty}^{\infty} P(x)x^2 dx - \langle X \rangle^2 = \langle X^2 \rangle - \langle X \rangle^2. \quad (1.4)$$

The *standard deviation* of  $X$ , denoted by  $\sigma_X$  and defined as the square root of the variance, is a measure of how broad the probability density for  $X$  is – that is, how much we can expect  $X$  to deviate from its mean value.

An important example of a probability density is the Gaussian, given by

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (1.5)$$

The mean of this Gaussian probability density is  $\langle X \rangle = \mu$  and the variance is  $V(x) = \sigma^2$ . A plot of this probability density is given in Figure 1.2.

## 1.2 Independence

Two random variables are referred to as being *independent* if neither of their probability densities depends on the value of the other variable. For example, if we rolled our six-sided die two times, and called the outcome of the first roll  $X$ , and the outcome of the second roll  $Y$ , then these two random variables would be independent. Further, we speak of the event  $X = 3$  (when the first die roll comes up as 3) and the event  $Y = 6$  as being independent. When two events are independent, the probability that both of them occur (that  $X = 3$  and  $Y = 6$ ) is the *product* of the probabilities that each occurs. One often states this by saying that “independent probabilities multiply”. The reason for this is fairly clear if we consider first making the die roll to obtain  $X$ . Only if  $X = 3$  do we then make the second roll, and only if that comes up 6 do we get the result  $X = 3$  and  $Y = 6$ . If the first roll only comes up 3 one eighth of the time, and the second comes up 6 one sixth of the time, then we will only get both of them  $1/8 \times 1/6 = 1/48$  of the time.

Once again this is just as true for independent random variables that take a continuum of values. In this case we speak of the “joint probability density”,  $P(x, y)$ , that  $X$  is equal to  $x$  and  $Y$  is equal to  $y$ . This joint probability density is the product of the probability densities for each of the two independent random variables, and we write this as  $P(x, y) = P_X(x)P_Y(y)$ . The probability that  $X$  falls within the interval  $[a, b]$  and  $Y$  falls in the interval  $[c, d]$  is then

$$\begin{aligned} \text{Prob}(X \in [a, b] \text{ and } Y \in [c, d]) &= \int_a^b \int_c^d P(x, y) dy dx \\ &= \int_a^b \int_c^d P_X(x)P_Y(y) dy dx = \left( \int_a^b P_X(x) dx \right) \left( \int_c^d P_Y(y) dy \right) \\ &= \text{Prob}(X \in [a, b]) \times \text{Prob}(Y \in [c, d]). \end{aligned}$$

In general, if we have a joint probability density,  $P(x_1, \dots, x_N)$ , for the  $N$  variables  $X_1, \dots, X_N$ , then the expectation value of a function of the variables,  $f(X_1, \dots, X_N)$ , is given by integrating the joint probability density over all the variables:

$$\langle f(X_1, \dots, X_N) \rangle = \int_{-\infty}^{\infty} f(x_1, \dots, x_N) P(x_1, \dots, x_N) dx_1 \dots dx_N. \quad (1.6)$$

It is also worth noting that when two variables are independent, then the expectation value of their product is simply the product of their individual expectation values. That is

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle. \quad (1.7)$$

## 1.3 Dependent random variables

Random variables,  $X$  and  $Y$ , are said to be *dependent* if their joint probability density,  $P(x, y)$ , does not factor into the product of their respective probability densities.

To obtain the probability density for one of the variables alone (say  $X$ ), we integrate the joint probability density over all values of the other variable (in this case  $Y$ ). This is because, for each value of  $X$ , we want to know the total probability summed over all the mutually exclusive values that  $Y$  can take. In this context, the probability densities for the single variables are referred to as the *marginals* of the joint density.

If we know nothing about the value of  $Y$ , then our probability density for  $X$  is just the marginal

$$P_X(x) = \int_{-\infty}^{\infty} P(x, y) dy. \quad (1.8)$$

If  $X$  and  $Y$  are dependent, and we learn the value of  $Y$ , then in general this will change our probability density for  $X$  (and vice versa). The probability density for  $X$  *given* that we know that  $Y = y$ , is written  $P(x|y)$ , and is referred to as the *conditional probability density for  $X$  given  $Y$* .

To see how to calculate this conditional probability, we note first that  $P(x, y)$  with  $y = a$  gives the *relative* probability for different values of  $x$  given that  $Y = a$ . To obtain the conditional probability density for  $X$  given that  $Y = a$ , all we have to do is divide  $P(x, a)$  by its integral over all values of  $x$ . This ensures that the integral of the conditional probability is 1. Since this is true for any value of  $y$ , we have

$$P(x|y) = \frac{P(x, y)}{\int_{-\infty}^{\infty} P(x, y) dx}. \quad (1.9)$$

Note also that since

$$P_Y(y) = \int_{-\infty}^{\infty} P(x, y) dx, \quad (1.10)$$

if we substitute this into the equation for the conditional probability above (Eq. (1.9)) we have

$$P(x|y) = \frac{P(x, y)}{P_Y(y)}, \quad (1.11)$$

and further that  $P(x, y) = P(x|y)P_Y(y)$ .

As an example of a conditional probability density consider a joint probability density for  $X$  and  $Y$ , where the probability density for  $Y$  is a Gaussian with zero

mean, and that for  $X$  is a Gaussian whose mean is given by the value of  $Y$ . In this case  $X$  and  $Y$  are not independent, and we have

$$P(x, y) = P(x|y)P(y) = \frac{e^{-(1/2)(x-y)^2}}{\sqrt{2\pi}} \times \frac{e^{-(1/2)y^2}}{\sqrt{2\pi}} = \frac{e^{-(1/2)(x-y)^2-(1/2)y^2}}{2\pi}, \quad (1.12)$$

where we have chosen the variance of  $Y$ , and of  $X$  given  $Y$  to be unity. Generally, when two random variables are dependent,  $\langle XY \rangle \neq \langle X \rangle \langle Y \rangle$ .

#### 1.4 Correlations and correlation coefficients

The expectation value of the product of two random variables is called the *correlation* of the two variables. The reason that we call this quantity a *correlation* is that, if the two random variables have zero mean and fixed variance, then the larger the value of the correlation, the more the variables tend to fluctuate *together* rather than independently; that is, if one is positive, then it is more likely that the other is positive. The value of the correlation therefore indicates how *correlated* the two variables are.

Of course, if we increase the variance of either of the two variables then the correlation will also increase. We can remove this dependence, and obtain a quantity that is a clearer indicator of the mutual dependence between the two variables by dividing the correlation by  $\sqrt{V(X)V(Y)}$ . This new quantity is called the correlation coefficient of  $X$  and  $Y$ , and is denoted by  $C_{XY}$ :

$$C_{XY} \equiv \frac{\langle XY \rangle}{\sqrt{V(X)V(Y)}}. \quad (1.13)$$

If the means of  $X$  and  $Y$  are not zero, then we can remove these when we calculate the correlation coefficient, so as to preserve its properties. Thus, in general, the correlation coefficient is defined as

$$C_{XY} \equiv \frac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\sqrt{V(X)V(Y)}} = \frac{\langle XY \rangle - \langle X \rangle \langle Y \rangle}{\sqrt{V(X)V(Y)}}. \quad (1.14)$$

The quantity on the top line,  $\langle XY \rangle - \langle X \rangle \langle Y \rangle$  is called the *covariance* of  $X$  and  $Y$ , and is zero if  $X$  and  $Y$  are independent. The correlation coefficient is therefore zero if  $X$  and  $Y$  are independent (completely uncorrelated), and is unity if  $X = cY$ , for some positive constant  $c$  (perfect correlation). If  $X = -cY$ , then the correlation coefficient is  $-1$ , and we say that the two variables are perfectly *anti-correlated*. The correlation coefficient provides a rough measure of the mutual dependence of two random variables, and one which is relatively easy to calculate.

#### 1.5 Adding random variables together

When we have two continuous random variables,  $X$  and  $Y$ , with probability densities  $P_X$  and  $P_Y$ , it is often useful to be able to calculate the probability density of the random variable whose value is the sum of them:  $Z = X + Y$ . It turns out that the probability density for  $Z$  is given by

$$P_Z(z) = \int_{-\infty}^{\infty} P_X(s - z)P_Y(s)ds \equiv P_X * P_Y, \quad (1.15)$$

which is called the *convolution* of  $P_X$  and  $P_Y$  [1]. Note that the convolution of two functions, denoted by “ $*$ ”, is another function. It is, in fact, quite easy to see directly why the above expression for  $P_Z(z)$  is true. For  $Z$  to equal  $z$ , then if  $Y = y$ ,  $X$  must be equal to  $z - y$ . The probability (density) for that to occur is  $P_Y(y)P_X(z - y)$ . To obtain the total probability (density) that  $Z = z$ , we need to sum this product over all possible values of  $Y$ , and this gives the expression for  $P_Z(z)$  above.

It will be useful to know the mean and variance of a random variable that is the sum of two or more random variables. It turns out that if  $X = X_1 + X_2$ , then the mean of  $X$  is

$$\langle X \rangle = \langle X_1 \rangle + \langle X_2 \rangle, \quad (1.16)$$

and if  $X_1$  and  $X_2$  are independent, then

$$V_X = V_{X_1} + V_{X_2}. \quad (1.17)$$

That is, when we add independent random variables both the means and variances add together to give the mean and variance of the new random variable. It follows that this remains true when we add any number of independent random variables together, so that, for example,  $\langle \sum_{n=1}^N X_n \rangle = \sum_{n=1}^N \langle X_n \rangle$ .

If you have ever taken an undergraduate physics lab, then you will be familiar with the notion that averaging the results of a number of independent measurements produces a more accurate result. This is because the variances of the different measurement results add together. If all the measurements are made using the same method, we can assume the results of all the measurements have the same mean,  $\mu$ , and variance,  $V$ . If we average the results,  $X_n$ , of  $N$  of these independent measurements, then the mean of the average is

$$\mu_{av} = \left\langle \sum_{n=1}^N \frac{X_n}{N} \right\rangle = \sum_{n=1}^N \frac{\mu}{N} = \mu. \quad (1.18)$$

But because we are dividing each of the variables by  $N$ , the variance of each goes down by  $1/N^2$ . Because it is the variances that add together, the variance of the

sum is

$$V_{av} = V \left[ \sum_{n=1}^N \frac{X_n}{N} \right] = \sum_{n=1}^N \frac{V}{N^2} = \frac{V}{N}. \quad (1.19)$$

Thus the variance gets smaller as we add more results together. Of course, it is not the variance that quantifies the uncertainty in the final value, but the standard deviation. The standard deviation of each measurement result is  $\sigma = \sqrt{V}$ , and hence the standard deviation of the average is

$$\sigma_{av} = \sqrt{\frac{V}{N}} = \frac{\sigma}{\sqrt{N}}. \quad (1.20)$$

The accuracy of the average therefore increases as the square root of the number of measurements.

## 1.6 Transformations of a random variable

If we know the probability density for a random variable  $X$ , then it can be useful to know how calculate the probability density for a random variable,  $Y$ , that is some function of  $X$ . This is referred to as a *transformation* of a random variable because we can think of the function as transforming  $X$  into a new variable  $Y$ . Let us begin with a particularly simple example, in which  $Y$  is a linear function of  $X$ . This means that  $Y = aX + b$  for some constants  $a$  and  $b$ . In this case it is not that hard to see the answer directly. Since we have multiplied  $X$  by  $a$ , the probability density will be stretched by a factor of  $a$ . Then adding  $b$  will shift the density by  $b$ . The result is that the density for  $Y$  is  $Q(Y) = P(y/a - b/a)/a$ .

To calculate the probability density for  $Y = aX + b$  in a more systematic way (which we can then use for much more general transformations of a random variable) we use the fact that the probability density for  $Y$  determines the average value of a function of  $Y$ ,  $f(Y)$ , through the relation

$$\langle f(Y) \rangle = \int_{-\infty}^{\infty} P(y) f(y) dy. \quad (1.21)$$

Now, since we know that  $Y = g(X) = aX + b$ , we also know that

$$\langle f(Y) \rangle = \int_{-\infty}^{\infty} P(x) f(y) dx = \int_{-\infty}^{\infty} P(x) f(ax + b) dx. \quad (1.22)$$

Changing variables in the integral from  $x$  to  $y$  we have

$$\langle f(Y) \rangle = \int_{-\infty}^{\infty} P(x) f(ax + b) dx = \frac{1}{a} \int_{-\infty}^{\infty} P(y/a - b/a) f(y) dy. \quad (1.23)$$

Thus the density for  $Y$  is

$$Q(y) = \frac{1}{a} P(y/a - b/a). \quad (1.24)$$

In addition, it is simple to verify that  $\langle Y \rangle = a\langle X \rangle + b$  and  $V_Y = a^2 V_X$ .

More generally, we can derive an expression for the probability density of  $Y$  when  $Y$  is an arbitrary function of a random variable. If  $Y = g(X)$ , then we determine the probability density for  $Y$  by changing variables in the same way as above. We begin by writing the expectation value of a function of  $Y$ ,  $f(Y)$ , in terms of  $P(x)$ . This gives

$$\langle f(Y) \rangle = \int_{x=a}^{x=b} P(x) f(g(x)) dx, \quad (1.25)$$

where  $a$  and  $b$  are, respectively, the lower and upper limits on the values that  $X$  can take. Now we transform this to an integral over the values of  $Y$ . Denoting the inverse of the function  $g$  as  $g^{-1}$ , so that  $X = g^{-1}(Y)$ , we have

$$\begin{aligned} \langle f(Y) \rangle &= \int_{x=a}^{x=b} P(x) f(g(x)) dx = \int_{y=g(a)}^{y=g(b)} P(g^{-1}(y)) \left( \frac{dx}{dy} \right) f(y) dy \\ &= \int_{y=g(a)}^{y=g(b)} \frac{P(g^{-1}(y))}{g'(x)} f(y) dy = \int_{y=g(a)}^{y=g(b)} \frac{P(g^{-1}(y))}{g'(g^{-1}(y))} f(y) dy. \end{aligned} \quad (1.26)$$

We now identify the function that multiplies  $f(y)$  inside the integral over  $y$  as the probability density for  $Y$ . But in doing so we have to be a little bit careful. If the lower limit for  $y$ ,  $g(a)$ , is *greater* than the upper limit for  $y$ , then the probability density we get will be negative to compensate for this inversion of the integral limits. So the probability density is actually the absolute value of the function inside the integral. The probability density for  $y$  is therefore

$$Q(y) = \frac{P(g^{-1}(y))}{|g'(g^{-1}(y))|}. \quad (1.27)$$

One must realize also that this expression for  $Q(y)$  only works for functions that map a single value of  $x$  to a single value of  $y$  (invertible functions), because in the change of variables in the integral we assumed that  $g$  was invertible. For non-invertible functions, for example  $y = x^2$ , one needs to do the transformation of the integral on a case-by-case basis to work out  $Q(y)$ .

### 1.7 The distribution function

The *probability distribution function*, which we will call  $D(x)$ , of a random variable  $X$  is defined as the probability that  $X$  is less than or equal to  $x$ . Thus

$$D(x) = \text{Prob}(X \leq x) = \int_{-\infty}^x P(z) dz. \quad (1.28)$$

In addition, the fundamental theorem of calculus tells us that

$$P(x) = \frac{dD(x)}{dx}. \quad (1.29)$$

### 1.8 The characteristic function

Another very useful definition is that of the *characteristic function*,  $\chi(s)$ . The characteristic function is defined as the *Fourier transform* of the probability density. Thus before we discuss the characteristic function, we need to explain what the Fourier transform is. The Fourier transform of a function  $P(x)$  is another function given by

$$\chi(s) = \int_{-\infty}^{\infty} P(x)e^{isx} dx. \quad (1.30)$$

The Fourier transform has many useful properties. One of them is the fact that it has a simple inverse, allowing one to perform a transformation on  $\chi(s)$  to get back  $P(x)$ . This inverse transform is

$$P(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(s)e^{-isx} ds. \quad (1.31)$$

Another very useful property is the following. If we have two functions  $F(x)$  and  $G(x)$ , then the Fourier transform of their convolution is simply the *product* of their respective Fourier transforms! This can be very useful because a product is always easy to calculate, but a convolution is not. Because the density for the sum of two random variables in the convolution of their respect densities, we now have an alternate way to find the probability density of the sum of two random variables: we can either convolve their two densities, or we can calculate the characteristic functions for each, multiply these together, and then take the inverse Fourier transform.

Showing that the Fourier transform of the convolution of two densities is the product of their respective Fourier transforms is not difficult, but we do need to use the Dirac  $\delta$ -function, denoted by  $\delta(x)$ . The Dirac  $\delta$ -function is zero everywhere except at  $t = 0$ , where it is infinite. It is defined in such a way that it integrates to

unity:

$$\int_{-\infty}^{\infty} \delta(x) dx = 1. \quad (1.32)$$

We get the  $\delta$ -function if we take the limit in which a function with fixed area becomes increasingly sharply peaked about  $x = 0$ . If we shift the  $\delta$ -function so that it is peaked at  $x = c$ , multiply it by another function,  $f(x)$ , and integrate over all space, this picks out the value of  $f(x)$  at  $x = c$ :

$$\int_{-\infty}^{\infty} \delta(x - c)f(x) dx = f(c). \quad (1.33)$$

The  $\delta$ -function can be rigorously defined using the theory of distributions (which was introduced for this purpose) or using measure theory. The  $\delta$ -function is very useful when using Fourier transforms. The  $\delta$ -function is the Fourier transform of the constant function  $f = 1/(2\pi)$ . That is

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{isx} dx = \delta(s). \quad (1.34)$$

With the above results we can now show that the Fourier transform of the convolution of two functions,  $P(x)$  and  $Q(x)$ , is the product of their respective Fourier transforms,  $\chi_P(s)$  and  $\chi_Q(s)$ . Denoting the convolution of  $P(x)$  and  $Q(x)$  as  $R(x) = P * Q$ , and using the definition of the Fourier transform, we have

$$\begin{aligned} \chi_R(s) &= \int_{-\infty}^{\infty} R(x)e^{isx} dx = \int \left[ \int P(y)Q(x-y) dy \right] e^{isx} dx \\ &= \frac{1}{(2\pi)^2} \iiint \chi_P(s')e^{-is'y} \chi_Q(s'')e^{-is''(x-y)} e^{isx} ds' ds'' dy dx \\ &= \frac{1}{(2\pi)^2} \iint \left[ \int e^{ix(s-s'')} dx \int e^{iy(s''-s')} dy \right] \chi_P(s') \chi_Q(s'') ds' ds'' \\ &= \iint \delta(s-s'') \delta(s''-s') \chi_P(s') \chi_Q(s'') ds' ds'' \\ &= \int \delta(s-s'') \chi_P(s'') \chi_Q(s'') ds'' = \chi_P(s) \chi_Q(s), \end{aligned} \quad (1.35)$$

where all the integrals are from  $-\infty$  to  $\infty$ .

One can also define the characteristic function for discrete random variables, and it has all the same properties. To do this, we again use the handy  $\delta$ -function. Let us say that we have a discrete random variable  $X$ , that takes the values  $\alpha_n$ , for  $n = 1, 2, \dots, N$ . If the probabilities for the values  $\alpha_n$  are  $P_n$ , then we can write a *probability density* for  $X$  using  $\delta$ -functions at the locations  $\alpha_n$ . This probability

density for  $X$  is

$$P(x) = \sum_{n=1}^N p_n \delta(x - \alpha_n). \quad (1.36)$$

The characteristic function for  $X$  is then the Fourier transform of this probability density:

$$\chi_X(s) = \int_{-\infty}^{\infty} \left[ \sum_{n=1}^N p_n \delta(x - \alpha_n) \right] e^{isx} dx = \sum_{n=1}^N \exp\{is\alpha_n\} p_n. \quad (1.37)$$

Note that regardless of whether the characteristic function is for a continuous or discrete random variable, one can always write it as the expectation value

$$\chi_X(s) = \langle e^{isX} \rangle, \quad (1.38)$$

where  $X$  is the random variable. For a probability density for a set of random variables,  $X_1, \dots, X_N$ , the characteristic function is

$$\chi_{\mathbf{X}}(\mathbf{s}) = \langle e^{i\mathbf{s} \cdot \mathbf{X}} \rangle, \quad (1.39)$$

where  $\mathbf{X} = (X_1, \dots, X_N)$  is the vector of random variables, and  $\mathbf{s} = (s_1, \dots, s_N)$ .

### 1.9 Moments and cumulants

For a random variable  $X$ , the expectation value of  $X^n$ ,  $\langle X^n \rangle$ , is called the  $n$ th *moment*. The moments can be calculated from the derivatives of the characteristic function, evaluated at  $s = 0$ . We can see this by expanding the characteristic function as a Taylor series:

$$\chi(s) = \sum_{n=0}^{\infty} \frac{\chi^{(n)}(0)s^n}{n!}, \quad (1.40)$$

where  $\chi^{(n)}(s)$  is the  $n$ th derivative of  $\chi(s)$ . But we also have

$$\chi(s) = \langle e^{isX} \rangle = \left\langle \sum_{n=0}^{\infty} \frac{(isX)^n}{n!} \right\rangle = \sum_{n=0}^{\infty} \frac{(i)^n \langle X^n \rangle s^n}{n!}. \quad (1.41)$$

Equating these two expressions for the characteristic function gives us

$$\langle X^n \rangle = \frac{\chi^{(n)}(0)}{(i)^n}. \quad (1.42)$$

The  $n$ th-order *cumulant* of  $X$ ,  $\kappa_n$ , which is a polynomial in the first  $n$  moments, is defined as the  $n$ th derivative of the *log* of the characteristic function, also evaluated at zero. The reason for this apparently rather odd definition is that it gives the

cumulants a special property. If we add two random variables together, then the  $n$ th cumulant of the result is the *sum* of the  $n$ th cumulants of the two random variables: when we add random variables, all the cumulants merely add together. The first cumulant is simply the mean, and the second cumulant is the variance. The next two cumulants are given by

$$\kappa_3 = \langle X^3 \rangle - 3\langle X^2 \rangle \langle X \rangle + 2\langle X \rangle^3, \quad (1.43)$$

$$\kappa_4 = \langle X^4 \rangle - 3\langle X^2 \rangle^2 - 4\langle X^3 \rangle \langle X \rangle + 12\langle X^2 \rangle \langle X \rangle^2 - 6\langle X \rangle^4. \quad (1.44)$$

Note that the Gaussian probability density is special, because all cumulants above second order vanish. This is because the characteristic function for the Gaussian is also a Gaussian. Taking the log cancels out the exponential in the Gaussian, and we are left with a quadratic in  $s$ , so that the Taylor series stops after  $n = 2$ .

For probability densities that contain more than one variable, say  $x$  and  $y$ , the moments are defined as  $\langle X^n Y^m \rangle$ . When  $m = 0$ , these are the moments for  $X$  alone. When  $n = m = 1$  this is the correlation of  $X$  and  $Y$ , as defined in Section 1.2 above.

### 1.10 The multivariate Gaussian

It is possible to have a probability density for  $N$  variables, in which the marginal densities for each of the variables are all Gaussian, and where all the variables may be correlated. Defining a column vector of  $N$  random variables,  $\mathbf{x} = (x_1, \dots, x_N)^T$ , the general form of this *multivariate Gaussian* is

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det[\Gamma]}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Gamma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]. \quad (1.45)$$

Here  $\boldsymbol{\mu}$  is the vector of the means of the random variables, and  $\boldsymbol{\Gamma}$  is the matrix of the covariances of the variables,

$$\boldsymbol{\Gamma} = \langle \mathbf{X} \mathbf{X}^T \rangle - \langle \mathbf{X} \rangle \langle \mathbf{X}^T \rangle = \langle \mathbf{X} \mathbf{X}^T \rangle - \boldsymbol{\mu} \boldsymbol{\mu}^T. \quad (1.46)$$

Note that the diagonal elements of  $\boldsymbol{\Gamma}$  are the variances of the individual variables.

The characteristic function for this multivariate Gaussian is

$$\begin{aligned} \chi(\mathbf{s}) &= \int_{-\infty}^{\infty} P(\mathbf{x}) \exp(is \cdot \mathbf{x}) dx_1 \dots dx_N \\ &= \exp(-\mathbf{s}^T \boldsymbol{\Gamma} \mathbf{s}) \exp(is \cdot \boldsymbol{\mu}), \end{aligned} \quad (1.47)$$

where  $\mathbf{s} \equiv (s_1, \dots, s_N)$ .

It is also useful to know that all the higher moments of a Gaussian can be written in terms of the means and covariances. Defining  $\Delta X \equiv X - \langle X \rangle$ , for a

one-dimensional Gaussian we have

$$\langle \Delta X^{2n} \rangle = \frac{(2n-1)!(V_X)^n}{2^{n-1}(n-1)!}, \quad (1.48)$$

$$\langle \Delta X^{2n-1} \rangle = 0, \quad (1.49)$$

for  $n = 1, 2, \dots$

### Further reading

A beautiful account of probability theory and its applications in inference, measurement, and estimation (all being essentially the same thing), is given in *Probability Theory: The Logic of Science* by E. T. Jaynes [1]. We also recommend the collection of Jaynes' works on the subject, entitled *E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics* [2]. Both Fourier transforms and distributions (such as the  $\delta$ -function, also known as the “unit impulse”) are discussed in most textbooks on signal processing. See for example the text *Linear Systems* by Sze Tan [3], and *Signals and Systems* by Alan Oppenheim and Alan Willsky [4]. A nice introduction to the theory of distributions is given in *The Theory of Distributions: A Nontechnical Introduction* by Ian Richards and Heekyung Youn [5]. The application of probability to information theory may be found in Shannon's classic book, *The Mathematical Theory of Communication* [6].

### Exercises

1. From the joint probability given in Eq. (1.12), calculate the marginal probability density for  $X$ ,  $P(x)$ . Also, calculate the conditional probability for  $Y$  given  $X$ ,  $P(y|x)$ .
2. From the joint probability given in Eq. (1.12), calculate  $\langle XY \rangle$ . From the marginals of  $P(x, y)$ ,  $P(x)$  and  $P(y)$ , obtain the expectation values  $\langle X \rangle$  and  $\langle Y \rangle$ .
3. If  $Y = aX + b$ , show that  $\langle Y \rangle = a\langle X \rangle + b$  and  $V_Y = a^2 V_X$ .
4. If  $X$  is a random variable with the Gaussian probability density

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (1.50)$$

then what is the probability density of  $Z = X^2$ ?

5. (i) Calculate the characteristic function of the Gaussian probability density given in Eq. (1.5). (ii) Use this characteristic function to calculate the probability density of the sum of two random variables each with this Gaussian density. Hint: you don't have to calculate the inverse transform from the definition; you work it out directly from the answer to (i).

6. Calculate the characteristic function for the probability density

$$P(x) = \frac{1}{\sqrt{8\pi a x}} e^{-\frac{x}{2a}}, \quad (1.51)$$

where  $x$  takes values in the interval  $(0, \infty)$ .

7. The random variables  $X$  and  $Y$  are independent, and both have a Gaussian probability density with zero mean and unit variance. (i) What is the joint probability density for  $X$  and  $Z = X + Y$ ? (ii) What is the conditional probability density for  $Y$  given  $Z$ ? Hint: use the fact that  $P(y|z) = P(z|y)P(y)/P(z)$ . (Note: this relationship is called *Bayes' theorem*, and is the cornerstone of measurement theory, also known as statistical inference [1].)
8. The random variable  $X$  can take any value in the range  $(0, \infty)$ . Find a probability density for  $X$  such that the probability density for  $Y = \alpha X$ , where  $\alpha$  is any positive number, is the same as the density for  $X$ . Note that the probability density you get is *not normalizable* (that is, its integral is infinite). This probability density is *scale invariant*, and has uses in statistical inference, even though it is not normalizable [1].
9. The two random variables  $X$  and  $Y$  have a joint probability density such that the point  $(X, Y)$  is uniformly distributed on the unit circle.
  - (i) What is the joint probability density for  $X$  and  $Y$ ?
  - (ii) Calculate  $\langle X \rangle$ ,  $\langle Y \rangle$  and  $\langle XY \rangle$ .
  - (iii) Are  $X$  and  $Y$  independent?

# 2

## Differential equations

### 2.1 Introduction

A differential equation is an equation that involves one or more of the derivatives of a function. Let us consider a simple physical example. Say we have a toy train on a straight track, and  $x$  is the position of the train along the track. If the train is moving then  $x$  will be a function of time, and so we write it as  $x(t)$ . If we apply a constant force of magnitude  $F$  to the train, then its acceleration, being the second derivative of  $x$ , is equal to  $F/m$ , where  $m$  is the train's mass. Thus we have the simple *differential equation*

$$\frac{d^2x}{dt^2} = \frac{F}{m}. \quad (2.1)$$

To find how  $x$  varies with time, we need to find the function  $x(t)$  that satisfies this equation. In this case it is very simple, since we can just integrate both sides of the equation twice with respect to  $t$ . This gives

$$x(t) = Ft^2/2 + at + b, \quad (2.2)$$

where  $a$  and  $b$  are the constants of integration. These constants are determined by the initial value of  $x$ , which we will denote by  $x_0 \equiv x(0)$ , and the initial value of  $dx/dt$ , which we denote by  $v_0$ . To determine  $b$  from  $x_0$  and  $v_0$ , one sets  $t = 0$  in Eq. (2.2). To determine  $a$  one differentiates both sides of Eq. (2.2) with respect to time, and then sets  $t = 0$ . The resulting solution is  $x(t) = Ft^2/2 + v_0t + x_0$ .

Now let's take a more non-trivial example. Let's say we have a metal ball hanging on the end of a spring. If we call the equilibrium position of the ball  $x = 0$ , then the force on the ball is equal to  $-kx$ , where  $k$  is a positive constant and  $x$  is the vertical deviation of the ball from the equilibrium position. Thus the differential equation describing the motion of the ball (often called the *equation of motion*) is

$$\frac{d^2x}{dt^2} = -\left(\frac{k}{m}\right)x. \quad (2.3)$$

We will show how to solve equations like this later. For now we merely note that one solution is  $x(t) = x(0) \cos(\omega t)$ , where  $\omega = \sqrt{k/m}$ . This shows that the ball oscillates up and down after being positioned away from the equilibrium point (in this case by a distance  $x(0)$ ). Both Eq. (2.1) and Eq. (2.3) are called *second-order* differential equations, because they contain the second derivative of  $x$ .

### 2.2 Vector differential equations

We can change a second-order differential equation for some variable  $x$  into a set of two differential equations that only contain first derivatives. To do this we introduce a second variable, and set this equal to the first derivative of  $x$ . Using as our example the differential equation for the ball on the spring, Eq. (2.3), we now define  $p(t) = mdx/dt$ , where  $m$  is the mass of the ball (so that  $p(t)$  is the momentum of the ball). We now have two first-order differential equations for the two functions  $x(t)$  and  $p(t)$ :

$$\frac{dx}{dt} = \frac{p}{m} \quad \text{and} \quad \frac{dp}{dt} = -\left(\frac{k}{m}\right)x. \quad (2.4)$$

The differential equation for  $x$  is simply the definition of  $p$ , and the differential equation for  $p$  is obtained by substituting the definition of  $p$  into the original second-order differential equation for  $x$  (Eq. (2.3)).

We can now write this set of first-order differential equations in the “vector form”,

$$\frac{d}{dt} \begin{pmatrix} x \\ p \end{pmatrix} \equiv \begin{pmatrix} dx/dt \\ dp/dt \end{pmatrix} = \begin{pmatrix} p/m \\ -kx/m \end{pmatrix} = \frac{1}{m} \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix}. \quad (2.5)$$

Defining  $\mathbf{x} = (x, p)^T$  and  $A$  as the matrix

$$A = \frac{1}{m} \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix}, \quad (2.6)$$

we can write the set of equations in the compact form

$$\dot{\mathbf{x}} \equiv \frac{d\mathbf{x}}{dt} = A\mathbf{x}. \quad (2.7)$$

If the elements of the matrix  $A$  do not depend on  $\mathbf{x}$ , as in the equation above, then this differential equation is referred to as a *linear* first-order vector differential equation.

We can use a similar procedure to transform any  $n$ th-order differential equation for  $x$  into a set of  $n$  first-order differential equations. In this case one defines  $n$  variables,  $x_1, \dots, x_n$ , with  $x_1 \equiv x$ , and  $x_m \equiv dx_{m-1}/dt$ , for  $m = 2, \dots, n$ . The definitions of the variables  $x_2$  to  $x_n$  give us  $n - 1$  differential equations, and the

final differential equation, being the equation for  $dx_n/dt = d^n x/dt^n$ , is given by substituting  $x_n$  into the original  $n$ th-order differential equation for  $x$ .

### 2.3 Writing differential equations using differentials

We now introduce another way of writing differential equations, as this will be most useful for the stochastic differential equations that we will encounter later. Instead of focussing on the derivative of  $x$  at each time  $t$ , we will instead consider the *change* in  $x$  in an infinitesimal time-step  $dt$ . We will call this change  $dx$ . By *infinitesimal* we mean a time-step that is small enough that only the first derivative of  $x$  contributes significantly to the change that  $x$  experiences in that interval. The change  $dx$  is given by

$$dx = \frac{dx}{dt} dt. \quad (2.8)$$

We can write our differential equations in terms of  $dx$  and  $dt$  instead of using the derivatives as we did above. Thus the differential equation given by Eq. (2.5) can be written instead as

$$d \begin{pmatrix} x \\ p \end{pmatrix} \equiv \begin{pmatrix} dx \\ dp \end{pmatrix} = \frac{1}{m} \begin{pmatrix} pdt \\ -kxdt \end{pmatrix} = \frac{1}{m} \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} dt, \quad (2.9)$$

or in the more compact form

$$\mathbf{dx} = A \mathbf{x} dt, \quad (2.10)$$

where  $\mathbf{x} = (x, p)^T$ , and  $A$  is as defined in Eq. (2.6). The infinitesimal increments  $dx$ ,  $dt$ , etc., are called “differentials”, and so writing differential equations in this way is often referred to as writing them in “differential form”.

### 2.4 Two methods for solving differential equations

The usual method of solving a first-order differential equation of the form

$$\frac{dx}{dt} = g(t)f(x) \quad (2.11)$$

is to divide by  $f(x)$ , multiply by  $dt$ , and then integrate both sides:

$$\int \frac{dx}{f(x)} = \int g(t)dt + C, \quad (2.12)$$

where  $C$  is the constant of integration. This constant is determined by the initial condition after one has performed the integration and solved the resulting equation

for  $x$ . This method is called “separation of variables”, because it works by separating out all the dependence on  $x$  to one side, and the dependence on  $t$  to the other. It works for linear first-order equations, as well as many nonlinear first-order equations. (An equation is linear if it contains the variables and their derivatives, but not higher powers or any more complex functions of these things.) We will need to use the above method to solve nonlinear equations in later chapters, but for now we are concerned only with the linear case.

We now present an alternative method for solving linear differential equations, because this will be useful when we come to solving stochastic equations, and helps to get us used to thinking in terms of differentials. Let’s say we have the simple linear differential equation

$$dx = -\gamma x dt. \quad (2.13)$$

This tells us that the value of  $x$  at time  $t + dt$  is the value at time  $t$  plus  $dx$ . That is

$$x(t + dt) = x(t) - \gamma x(t)dt = (1 - \gamma dt)x(t). \quad (2.14)$$

To solve this we note that to first order in  $dt$  (that is, when  $dt$  is very small)  $e^{\alpha dt} \approx 1 + \alpha dt$ . We can therefore write the equation for  $x(t + dt)$  as

$$x(t + dt) = x(t) - \gamma x(t)dt = e^{-\gamma dt}x(t). \quad (2.15)$$

This tells us that to move  $x$  from time  $t$  to  $t + dt$  we merely have to multiply  $x(t)$  by the factor  $e^{-\gamma dt}$ . So to move by two lots of  $dt$  we simply multiply by this factor twice:

$$x(t + 2dt) = e^{-\gamma dt}x(t + dt) = e^{-\gamma dt}[e^{-\gamma dt}x(t)] = e^{-2\gamma dt}x(t). \quad (2.16)$$

To obtain  $x(t + \tau)$  all we have to do is apply this relation repeatedly. Let us say that  $dt = \tau/N$  for  $N$  as large as we want. Thus  $dt$  is a small but finite time-step, and we can make it as small as we want. That means that to evolve  $x$  from time  $t$  to  $t + \tau$  we have to apply Eq. (2.15)  $N$  times. Thus

$$x(t + \tau) = (e^{-\gamma dt})^N x(t) = e^{-\gamma \sum_{n=1}^N dt} x(t) = e^{-\gamma N dt} x(t) = e^{-\gamma \tau} x(t) \quad (2.17)$$

is the solution to the differential equation. If  $\gamma$  is a function of time, so that the equation becomes

$$dx = -\gamma(t)x dt, \quad (2.18)$$

we can still use the above technique. As before we set  $dt = \tau/N$  so that it is a small finite time-step. But this time we have to explicitly take the limit as  $N \rightarrow \infty$

to obtain the solution to the differential equation:

$$\begin{aligned} x(t + \tau) &= \lim_{N \rightarrow \infty} \prod_{n=1}^N e^{-\gamma(t+n\Delta t)dt} x(t) \\ &= \lim_{N \rightarrow \infty} e^{-\sum_{n=1}^N \gamma(t+n\Delta t)dt} x(t) \\ &= e^{-\int_t^{t+\tau} \gamma(s)ds} x(t). \end{aligned} \quad (2.19)$$

The equation we have just solved is the simplest *linear* differential equation. All linear differential equations can be written in the form of Eq. (2.10) above, where the matrix  $A$  is independent of  $\mathbf{x}$ .

#### 2.4.1 A linear differential equation with driving

We will shortly show how one solves linear differential equations when there is more than one variable. But before we do, we consider the simple linear differential equation given by Eq. (2.13) with the addition of a *driving term*. A driving term is a function of time that is independent of the variable. So the single-variable linear equation with driving is

$$\frac{dx}{dt} = -\gamma x + f(t), \quad (2.20)$$

where  $f$  is any function of time. To solve this we first transform to a new variable,  $y(t)$ , defined as

$$y(t) = x(t)e^{\gamma t}. \quad (2.21)$$

Note that  $y(t)$  is defined precisely so that if  $x(t)$  was a solution to Eq. (2.13), then  $y$  would be constant. Now we calculate the differential equation for  $y$ . This is

$$\frac{dy}{dt} = \left( \frac{\partial y}{\partial x} \right) \frac{dx}{dt} + \frac{\partial y}{\partial t} = e^{\gamma t} f(t). \quad (2.22)$$

The equation for  $y$  is solved merely by integrating both sides, and the solution is

$$y(t) = y_0 + \int_0^t e^{\gamma s} f(s)ds, \quad (2.23)$$

where we have defined  $y_0$  as the value of  $y$  at time  $t = 0$ . Now we can easily obtain  $x(t)$  from  $y(t)$  by inverting Eq. (2.21). This gives us the solution to Eq. (2.20), which is

$$x(t) = x_0 e^{-\gamma t} + e^{-\gamma t} \int_0^t e^{\gamma s} f(s)ds = x_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-s)} f(s)ds. \quad (2.24)$$

We can just as easily solve a linear equation when the coefficient  $\gamma$  is a function of time. In this case we transform to  $y(t) = x(t) \exp[\Gamma(t)]$ , where

$$\Gamma(t) \equiv \int_0^t \gamma(s)ds, \quad (2.25)$$

and the solution is

$$x(t) = x_0 e^{-\Gamma(t)} + e^{-\Gamma(t)} \int_0^t e^{\Gamma(s)} f(s)ds. \quad (2.26)$$

## 2.5 Solving vector linear differential equations

We can usually solve a linear equation with more than one variable,

$$\dot{\mathbf{x}} = A\mathbf{x}, \quad (2.27)$$

by transforming to a new set of variables,  $\mathbf{y} = U\mathbf{x}$ , where  $U$  is a matrix chosen so that the equations for the new variables are *decoupled* from each other. That is, the equation for the vector  $\mathbf{y}$  is

$$\dot{\mathbf{y}} = D\mathbf{y}, \quad (2.28)$$

where  $D$  is a diagonal matrix. For many square matrices  $A$ , there exists a matrix  $U$  so that  $D$  is diagonal. This is only actually guaranteed if  $A$  is normal, which means that  $A^\dagger A = AA^\dagger$ . Here  $A^\dagger$  is the *Hermitian conjugate* of  $A$ , defined as the transpose of the complex conjugate of  $A$ . If there is no  $U$  that gives a diagonal  $D$ , then one must solve the differential equation using Laplace transforms instead, a method that is described in most textbooks on differential equations (for example [7]). If  $U$  exists then it is *unitary*, which means that  $U^\dagger U = UU^\dagger = I$ . The diagonal elements of  $D$  are called the *eigenvalues* of  $A$ .

There are systematic numerical methods to find the  $U$  and corresponding  $D$  for a given  $A$ , and numerical software such as Matlab and Mathematica include routines to do this. If  $A$  is two-by-two or three-by-three, then one can calculate  $U$  and  $D$  analytically, and we show how this is done in the next section.

If  $D$  is diagonal, so that

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \lambda_N \end{pmatrix}, \quad (2.29)$$

then for each element of  $\mathbf{y}$  (each variable),  $y_n$ , we have the simple equation

$$\dot{y}_n = \lambda_n y_n, \quad (2.30)$$

and this has the solution  $y_n(t) = y_n(0)e^{\lambda_n t}$ . The solution for  $\mathbf{y}$  is thus

$$\mathbf{y}(t) = \begin{pmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & e^{\lambda_N t} \end{pmatrix} \mathbf{y}(0) \equiv e^{Dt} \mathbf{y}(0), \quad (2.31)$$

where we have defined the exponential of a diagonal matrix,  $e^{Dt}$ .

To get the solution for  $\mathbf{x}(t)$ , we use the fact that  $U^\dagger U = I$ , from which it follows immediately that  $\mathbf{x} = U^\dagger \mathbf{y}$ . Substituting this, along with the definition of  $\mathbf{y}$  into the solution for  $\mathbf{y}$ , we get

$$\mathbf{x}(t) = U^\dagger e^{Dt} U \mathbf{x}(0). \quad (2.32)$$

Further, it makes sense to define the exponential of any square matrix  $A$  as

$$e^{At} = U^\dagger e^{Dt} U. \quad (2.33)$$

To see why, first note that, by substituting  $\mathbf{x}$  into the differential equation for  $\mathbf{y}$ , we find that

$$\dot{\mathbf{x}} = U^\dagger D U \mathbf{x}, \quad (2.34)$$

and thus  $At = U^\dagger Dt U$ . Because of this, the power series

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(At)^n}{n!} &= 1 + At + \frac{(At)^2}{2} + \frac{(At)^3}{3!} + \cdots \\ &= 1 + U^\dagger Dt U + \frac{U^\dagger Dt U U^\dagger Dt U}{2} + \frac{(U^\dagger Dt U)^3}{3!} + \cdots \\ &= 1 + U^\dagger Dt U + \frac{U^\dagger (Dt)^2 U}{2} + \frac{U^\dagger (Dt)^3 U}{3!} + \cdots \\ &= U^\dagger \left( 1 + Dt + \frac{(Dt)^2}{2} + \frac{(Dt)^3}{3!} + \cdots \right) U \\ &= U^\dagger e^{Dt} U. \end{aligned} \quad (2.35)$$

So  $U^\dagger e^{Dt} U$  corresponds precisely to the power series  $\sum_n^n (At)^n / n!$ , which is the natural generalization of the exponential function for a matrix  $At$ . Since the above relationship holds for all power series, the natural definition of *any* function of a square matrix  $A$  is

$$f(A) \equiv U^\dagger f(D) U \equiv U^\dagger \begin{pmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & f(\lambda_N) \end{pmatrix} U. \quad (2.36)$$

To summarize the above results, the solution to the vector differential equation

$$\dot{\mathbf{x}} = A \mathbf{x}, \quad (2.37)$$

is

$$\mathbf{x}(t) = e^{At} \mathbf{x}(0), \quad (2.38)$$

where

$$e^{At} \equiv \sum_n \frac{(At)^n}{n!} = U^\dagger e^{Dt} U. \quad (2.39)$$

We can now also solve any linear vector differential equation with driving, just as we did for the single-variable linear equation above. The solution to

$$\dot{\mathbf{x}} = A \mathbf{x} + \mathbf{f}(t), \quad (2.40)$$

where  $\mathbf{f}$  is now a vector of driving terms, is

$$\mathbf{x}(t) = e^{At} \mathbf{x}(0) + \int_0^t e^{A(t-s)} \mathbf{f}(s) ds. \quad (2.41)$$

## 2.6 Diagonalizing a matrix

To complete this chapter, we now show how to obtain the matrices  $U$  and  $D$  for a square  $N$  dimensional ( $N$  by  $N$ ) matrix  $A$ . It is feasible to use this to obtain analytical expressions for  $U$  and  $D$  when  $A$  is two dimensional, and for three dimensions if  $A$  has a sufficiently simple form.

First we need to find a set of  $N$  special vectors, called the *eigenvectors* of  $A$ . An eigenvector is a vector,  $\mathbf{v}$ , for which

$$A \mathbf{v} = \lambda \mathbf{v}, \quad (2.42)$$

where  $\lambda$  is a number (real or complex). To find all the eigenvectors we note that if Eq. (2.42) is true, then

$$(A - \lambda I) \mathbf{v} = 0, \quad (2.43)$$

where  $I$  is the  $N$ -dimensional identity matrix. This is true if and only if the determinant of  $A - \lambda I$  is zero. The equation

$$\det |A - \lambda I| = 0 \quad (2.44)$$

is an  $N$ th-order polynomial equation for  $\lambda$ . This has  $N$  solutions, giving in general  $N$  different eigenvalues. Note that some of the eigenvalues may be the same, because the polynomial may have repeated roots. In this case there are fewer than

$N$  distinct eigenvalues, but so long as  $A$  is Hermitian symmetric there will always be  $N$  distinct (and orthogonal) eigenvectors; each of the repeated roots will have more than one corresponding eigenvector. For  $N = 2$  and  $N = 3$  there exist analytical expressions for the roots of the polynomial, and thus for the eigenvalues. We will denote the  $N$  eigenvalues as  $\lambda_i$ ,  $i = 1, \dots, N$  (some of which may be repeated), and the corresponding eigenvectors as  $\mathbf{v}_i$ .

For each distinct (unrepeated) eigenvalue,  $\lambda_i$ , one determines the corresponding eigenvector by solving the equation

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad (2.45)$$

for  $\mathbf{v}_i$ . For an eigenvalue that is repeated  $m$  times (also known as a *degenerate* eigenvalue), we solve the same equation, except that now the solution is an  $m$ -dimensional vector space. In this case we then choose  $m$ -linearly independent vectors in this space. On obtaining these vectors, the Gramm–Schmidt [8] orthogonalization procedure can then be used to obtain  $m$  mutually orthogonal vectors that span the space. (If you are not familiar with vector spaces and the associated terminology, then this information can be obtained from a textbook on linear algebra [7].)

Having followed the above procedure, we now have  $N$  mutually orthogonal eigenvectors,  $\mathbf{v}_n$ , each with a corresponding eigenvalue,  $\lambda_n$ . All that remains to be done is to divide each eigenvector by its norm,  $|\mathbf{v}_n| = \sqrt{\mathbf{v}_n \cdot \mathbf{v}_n}$ , which gives us a set of orthonormal eigenvectors,  $\tilde{\mathbf{v}}_n = \mathbf{v}_n / |\mathbf{v}_n|$ . Defining  $U^\dagger$  by

$$U^\dagger = (\tilde{\mathbf{v}}_1 \tilde{\mathbf{v}}_2 \dots \tilde{\mathbf{v}}_N) \quad (2.46)$$

(so that the columns of  $U^\dagger$  are the orthonormal eigenvectors), and  $D$  by Eq. (2.29), it is straightforward to verify that  $UU^\dagger = I$  and that  $D = UAU^\dagger$ . It is also true, though not as obvious, that  $U^\dagger U = I$ .

### Exercises

1. The equation for the damped harmonic oscillator is:

$$m \frac{d^2x}{dt^2} + 2\gamma m \frac{dx}{dt} + kx = 0. \quad (2.47)$$

Assume that  $\gamma < \sqrt{k/m}$ , in the which the resonator is said to be “under-damped”. Show that  $x(t) = e^{-\gamma t} \cos(\omega t)x(0)$  is a solution to Eq. (2.47), and find the expression for the frequency  $\omega$  in terms of  $m$ ,  $k$  and  $\gamma$ .

2. Write Eq. (2.47) as a first-order vector differential equation.  
3. Solve the linear vector differential equation

$$\dot{\mathbf{x}} = A\mathbf{x} \quad (2.48)$$

with

$$A = \begin{pmatrix} 0 & \omega \\ -\omega & -\gamma \end{pmatrix}. \quad (2.49)$$

Do this by calculating the eigenvectors and eigenvalues of the matrix  $A$ . Assume that  $\gamma < \omega$ .

4. Use the answer to Exercise 3 above to write down the solution to the differential equation

$$\frac{d^2x}{dt^2} - \gamma \frac{dx}{dt} + \omega^2 x = f(t). \quad (2.50)$$

5. Calculate the eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 2 & 0 & \sqrt{\frac{2}{3}} \\ 0 & 2 & \frac{-1}{\sqrt{3}} \\ \sqrt{\frac{2}{3}} & \frac{-1}{\sqrt{3}} & 2 \end{pmatrix} \quad (2.51)$$

and use them to solve the differential equation

$$\dot{\mathbf{x}} = A\mathbf{x}. \quad (2.52)$$

6. Use the power series for the exponential function to show that

$$e^{\lambda\sigma} = \cos(\lambda)I + \sin(\lambda)\sigma, \quad (2.53)$$

where

$$\sigma = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (2.54)$$

and  $I$  is the two-by-two identity matrix.

# 3

## Stochastic equations with Gaussian noise

### 3.1 Introduction

We have seen that a differential equation for  $x$  is an equation that tells us how  $x$  changes in each infinitesimal time-step  $dt$ . In the differential equations we have considered so far, this increment in  $x$  is deterministic – that is, it is completely determined at each time by the differential equation. Now we are going to consider a situation in which this increment is not completely determined, but is a random variable. This is the subject of *stochastic differential equations*.

It will make things clearest if we begin by considering an equation in which time is divided into finite (that is, not infinitesimal) chunks of duration  $\Delta t$ . In this case  $x$  gets a finite increment,  $\Delta x$ , during each time-chunk. Such an equation is referred to as a *discrete-time* equation, or a *difference* equation, because it specifies the difference,  $\Delta x$ , between  $x(t)$  and  $x(t + \Delta t)$ . Since time moves in steps of  $\Delta t$ ,  $x$  takes values only at the set of times  $t_n = n\Delta t$ , where  $n = 0, 1, \dots, \infty$ .

Let us consider a discrete-time equation in which the change in  $x$ ,  $\Delta x$ , is the sum of a term that is proportional to  $x$ , and a “driving term” that is independent of  $x$ . We will make this driving term a function of time, thus giving it the freedom to be different at different times. The difference equation for  $x$  is

$$\Delta x(t_n) = x(t_n)\Delta t + f(t_n)\Delta t. \quad (3.1)$$

where  $f(t_n)\Delta t$  is the driving term.

Given the value of  $x$  at time  $t_n$ , the value at time  $t_{n+1}$  is then

$$x(t_n + \Delta t) = x(t_n) + \Delta x(t_n) = x(t_n) + x(t_n)\Delta t + f(t_n)\Delta t. \quad (3.2)$$

If we know the value of  $x$  at  $t = 0$ , then

$$x(\Delta t) = x(0)(1 + \Delta t) + f(0)\Delta t. \quad (3.3)$$

Now, what we are really interested in is what happens if the driving term,  $f(t_n)\Delta t$  is *random* at each time  $t_n$ ? This means replacing  $f(t_n)$  with a random variable,  $y_n$ , at each time  $t_n$ . Now the difference equation for  $x$  becomes

$$\Delta x(t_n) = x(t_n)\Delta t + y_n\Delta t, \quad (3.4)$$

and this is called a “stochastic difference equation”. This equation says that at each time  $t_n$  we pick a value for the random variable  $y_n$  (sampled from its probability density), and add  $y_n\Delta t$  to  $x(t_n)$ . This means that we can no longer predict exactly what  $x$  will be at some future time,  $T$ , until we arrive at that time, and all the values of the random increments up until  $T$  have been determined.

The solution for  $x$  at time  $\Delta t$  is

$$x(\Delta t) = x(0)(1 + \Delta t) + y_0\Delta t. \quad (3.5)$$

So  $x(\Delta t)$  is now a random variable. If  $x(0)$  is fixed (that is, not random), then  $x(\Delta t)$  is just a linear transformation of the random variable  $y_0$ . If  $x(0)$  is also random, then  $x(\Delta t)$  is a linear combination of the two random variables  $x(0)$  and  $y_0$ . Similarly, if we go to the next time-step, and calculate  $x(2\Delta t)$ , then this is a function of  $x(0)$ ,  $y_0$  and  $y_1$ . We see that at each time then, the solution to a stochastic difference equation,  $x(t_n)$ , is a random variable, and this random variable changes at each time-step. To solve a stochastic difference equation, we must therefore determine the *probability density* for  $x$  at all future times. Since  $x(t_n)$  is a function of all the noise increments  $y_n$ , as well as  $x(0)$ , this means calculating the probability density for  $x$  from the probability densities for the noise increments (and from the probability density for  $x(0)$  if it is also random). That is why we discussed summing random variables, and making transformations of random variables in Chapter 1.

Stochastic *differential* equations are obtained by taking the limit  $\Delta t \rightarrow 0$  of stochastic difference equations. Thus the solution of a stochastic differential equation is also a probability density for the value of  $x$  at each future time  $t$ . Just as in the case of ordinary (deterministic) differential equations, it is not always possible to find a closed-form expression for the solution of a stochastic differential equation. But for a number of simple cases it is possible. For stochastic equations that cannot be solved analytically, we can solve them numerically using a computer, and we describe this later in Chapter 6.

The reason that differential equations that are driven by random increments at each time-step are called *stochastic* differential equations is because a process that fluctuates randomly in time is called a *stochastic process*. Stochastic differential equations (SDEs) are thus driven by stochastic processes. The random increments that drive an SDE are also referred to as *noise*.

In addition to obtaining the probability density for  $x$  at the future times  $t_n$ , we can also ask how  $x$  evolves with time given a specific set of values for the random

increments  $y_n$ . A set of values of the random increments (values sampled from their probability densities) is called a *realization* of the noise. A particular evolution for  $x$  given a specific noise realization is called a *sample path* for  $x$ . In addition to wanting to know the probability density for  $x$  at future times, it can also be useful to know what properties the sample paths of  $x$  have. The full solution to an SDE is therefore really the complete set of all possible sample paths, and the probabilities for all these paths, but we don't usually need to know all this information. Usually all we need to know is the probability density for  $x$  at each time, and how  $x$  at one time is correlated with  $x$  at another time. How we calculate the latter will be discussed in Chapter 4. Note that since the solution to an SDE,  $x(t)$ , varies randomly in time, it is a stochastic process. If one wishes, one can therefore view an SDE as something that takes as an input one stochastic process (the driving noise), and produces another.

### 3.2 Gaussian increments and the continuum limit

In this chapter we are going to study stochastic differential equations driven by Gaussian noise. By “Gaussian noise” we mean that each of the random increments has a Gaussian probability density. First we consider the simplest stochastic difference equation, in which the increment of  $x$ ,  $\Delta x$ , consists solely of the random increment  $y_n \Delta t$ . Since Gaussian noise is usually called *Wiener noise*, we will call the random increment  $\Delta W_n = y_n \Delta t$ . The discrete differential equation for  $x$  is thus

$$\Delta x(t_n) = \Delta W_n. \quad (3.6)$$

Each *Wiener* increment is completely independent of all the others, and all have the same probability density, given by

$$P(\Delta W) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\Delta W)^2/(2\sigma^2)}. \quad (3.7)$$

This density is a Gaussian with zero mean, and we choose the variance to be  $\sigma^2 = V = \Delta t$ . This choice for the variance of the Wiener increment is very important, and we will see why shortly. We will often denote a Wiener increment in some time-step  $\Delta t$  simply as  $\Delta W$ , without reference to the subscript  $n$ . This notation is convenient because all the random increments have the same probability density, and even though independent of each other, are in that sense identical.

We can easily solve the difference equation for  $x$  simply by starting with  $x(0) = 0$ , and repeatedly adding  $\Delta x$ . This gives the solution

$$x_n \equiv x(n\Delta t) = \sum_{i=0}^{n-1} \Delta W_i. \quad (3.8)$$

So now we need to calculate the probability density for  $x_n$ . Since we know from Exercise 5 in Chapter 1 that the sum of two Gaussian random variables is also a Gaussian, we know that the probability density for  $x_n$  is Gaussian. We also know that the mean and variance of  $x_n$  is the sum of the means and variances of the  $\Delta W_i$ , because the  $\Delta W_i$  are all independent (see Section 1.5). We therefore have

$$\langle x_n \rangle = 0, \quad (3.9)$$

$$V(x_n) = n\Delta t, \quad (3.10)$$

and so

$$P(x_n) = \frac{1}{\sqrt{2\pi V}} e^{-x_n^2/(2V)} = \frac{1}{\sqrt{2\pi n\Delta t}} e^{-x_n^2/(2n\Delta t)}. \quad (3.11)$$

We now need to move from difference equations to *differential* equations. To do so we will consider solving the difference equation above, Eq. (3.6), for a given future time  $T$ , with  $N$  discrete time-steps. (So each time-step is  $\Delta t = T/N$ .) We then take the limit as  $N \rightarrow \infty$ . Proceeding in the same way as above, the solution  $x(T)$  is now

$$x(T) = \lim_{N \rightarrow \infty} \sum_{i=0}^{N-1} \Delta W_i \equiv \int_0^T dW(t) \equiv W(T). \quad (3.12)$$

Here we define a *stochastic integral*,  $W(T) = \int_0^T dW(t)$ , as the limit of the sum of all the increments of the Wiener process. A stochastic integral, being the sum of a bunch of random variables, is therefore a random variable. The important thing is that in many cases we can calculate the probability density for this random variable. We can easily do this in the case above because we know that the probability density for  $x(T)$  is Gaussian (since it is merely the sum of many independent Gaussian variables). Because of this, we need only to calculate its mean and variance. The mean is zero, because all the random variables in the sum have zero mean. To calculate the variance of  $x(T)$  it turns out that we don't even need to take the limit as  $N \rightarrow \infty$  because  $N$  factors out of the expression:

$$V(x(T)) = \sum_{i=0}^{N-1} V[\Delta W_i] = \sum_{i=0}^{N-1} \Delta t = N\Delta t = N \left( \frac{T}{N} \right) = T. \quad (3.13)$$

The probability density for  $W(T)$  is therefore

$$P(W(T)) = P(x(T)) \equiv P(x, T) = \frac{1}{\sqrt{2\pi T}} e^{-x^2/(2T)}. \quad (3.14)$$

Note that when writing the integral over the Wiener increments, we have included  $t$  explicitly as an argument to  $dW$ , just to indicate that  $dW$  changes with time. We

will often drop the explicit dependence of  $dW$  on  $t$ , and just write the stochastic integral as

$$W(T) = \int_0^T dW. \quad (3.15)$$

While we often loosely refer to the increments  $dW$  as the “Wiener process”, the Wiener process is actually defined as  $W(T)$ , and  $dW$  is, strictly speaking, an increment of the Wiener process.

The fact that the variance of  $x(T)$  is proportional to  $T$  is a result of the fact that we chose each of the Wiener increments to have variance  $\Delta t$ . Since there is one Wiener increment in each time step  $\Delta t$ , the variance of  $x$  grows by precisely  $\Delta t$  in each interval  $\Delta t$ , and is thus proportional to  $t$ . So what would happen if we chose  $V[\Delta W(\Delta t)]$  to be some other power of  $\Delta t$ ? To find this out we set  $V[\Delta W(\Delta t)] = \Delta t^\alpha$  and calculate once again the variance of  $x(T)$  (before taking the limit as  $N \rightarrow \infty$ ). This gives

$$V(x(T)) = \sum_{i=0}^{N-1} V[\Delta W_i] = N(\Delta t)^\alpha = N \left( \frac{T}{N} \right)^\alpha = N^{(1-\alpha)} T^\alpha. \quad (3.16)$$

Now we take the continuum limit  $N \rightarrow \infty$  so as to obtain a stochastic differential equation. When  $\alpha$  is greater than one we have

$$\lim_{N \rightarrow \infty} V(x(T)) = T^\alpha \lim_{N \rightarrow \infty} N^{(1-\alpha)} = 0, \quad (3.17)$$

and when  $\alpha$  is less than one we have

$$\lim_{N \rightarrow \infty} V(x(T)) = T^\alpha \lim_{N \rightarrow \infty} N^{(1-\alpha)} \rightarrow \infty. \quad (3.18)$$

Neither of these make sense for the purposes of obtaining a stochastic differential equation that describes real systems driven by noise. Thus we are *forced* to choose  $\alpha = 1$  and hence  $V(\Delta W_n) \propto \Delta t$ .

When we are working in the continuum limit the Gaussian increments,  $dW$ , are referred to as being *infinitesimal*. A general SDE for a single variable  $x(t)$  is then written as

$$dx = f(x, t)dt + g(x, t)dW. \quad (3.19)$$

Since the variance of  $dW$  must be proportional to  $dt$ , and since any constant of proportionality can always be absorbed into  $g(x, t)$ , the variance of  $dW$  is defined to be *equal* to  $dt$ . We can therefore write the probability density for  $dW$  as

$$P(dW) = \frac{e^{-(dW)^2/(2dt)}}{\sqrt{2\pi dt}}. \quad (3.20)$$

To summarize the main result of this section: if we have an SDE driven by an infinitesimal increment  $dW$  in each infinitesimal interval  $dt$ , then if all the increments are **Gaussian**, **independent** and **identical**, they *must* have variance proportional to  $dt$ .

### 3.3 Interlude: why Gaussian noise?

In this chapter we consider only noise that is Gaussian. This kind of noise is important because it is very common in physical systems. The reason for this is a result known as the *central limit theorem*. The central limit theorem says that if one sums together many independent random variables, then the probability density of the sum will be close to a Gaussian. As the number of variables in the sum tends to infinity, the resulting probability density tends exactly to a Gaussian. The only condition on the random variables is they all have a finite variance.

Now consider noise in a physical system. This noise is usually the result of many random events happening at the microscopic level. This could be the impacts of individual molecules, the electric force from many electrons moving randomly in a conductor, or the thermal jiggling of atoms in a solid. The total force applied by these microscopic particles is the sum of the random forces applied by each. Because the total force is the sum over many random variables, it has a Gaussian probability density. Since the microscopic fluctuations are usually fast compared to the motion of the system, we can model the noise as having a Gaussian probability density in each time-step  $\Delta t$ , where  $\Delta t$  is small compared to the time-scale on which the system moves. In fact, assuming that the noise increments are completely independent of each other from one infinitesimal time-step to the next is really an idealization (an approximation) which is not true in practice. Nevertheless, this approximation works very well, and we will explain why at the end of Section 4.6.

From a mathematical point of view, simple noise processes in which the random increment in each time interval  $dt$  is independent of all the previous random increments will usually be Gaussian for the same reason. This is because the random increment in each small but finite time interval  $\Delta t$  is the sum over the infinite number of increments for the infinitesimal intervals  $dt$  that make up that finite interval. There are two exceptions to this. One is processes in which the random increment in an infinitesimal time-step  $dt$  is not necessarily infinitesimal. The sample paths of such processes make instant and discrete jumps from time to time, and are thus not continuous. These are called *jump* or *point* processes, and we consider them in Chapter 8. Jump processes are actually quite common and have many applications. The other exception, which is much rarer in nature, happens when the noise increments remain infinitesimal, as in

Gaussian noise, but are drawn from a probability density with an infinite variance (one that avoids the central limit theorem). These processes will be discussed in Chapter 9.

### 3.4 Ito calculus

We saw in the previous section that the increments of Wiener noise,  $dW$ , have a variance proportional to  $dt$ . In this section we are going to discover a very surprising consequence of this fact, and the most unusual aspect of Wiener noise and stochastic equations. To set the stage, consider how we obtained the solution of the simple differential equation

$$dx = -\gamma x dt. \quad (3.21)$$

We solved this equation in Section 2.4 using the relation  $e^{\alpha dt} \approx 1 + \alpha dt$ . This relation is true because  $dt$  is a differential – that is, it is understood that when calculating the solution,  $x(t)$ , we will always take the limit in which  $dt \rightarrow 0$ , just as we did in solving the simple stochastic equation in the previous section (in fact, we can regard the use of the symbol  $dt$  as a shorthand notation for the fact that this limit will be taken). The approximation  $e^{\alpha dt} \approx 1 + \alpha dt$  works because the terms in the power series expansion for  $e^{\alpha dt}$  that are second-order or higher in  $dt$  ( $dt^2$ ,  $dt^3$ , etc.) will vanish in comparison to  $dt$  as  $dt \rightarrow 0$ .

The result of being able to ignore terms that are second-order and higher in the infinitesimal increment leads to the usual rules for differential equations. (It also means that any equation we write in terms of differentials  $dx$  and  $dt$  can alternatively be written in terms of derivatives.) However, we will now show that the second power of the stochastic differential  $dW$  does not vanish with respect to  $dt$ , and we must therefore learn a new rule for the manipulation of stochastic differential equations. It will also mean that we cannot write stochastic differential equations in terms of derivatives – we must use differentials.

Solving a differential equation involves summing the infinitesimal increments over all the time-steps  $dt$ . To examine whether  $(dW)^2$  makes a non-zero contribution to the solution, we must therefore sum  $(dW)^2$  over all the time-steps for a finite time  $T$ . To do this we will return to a discrete description so that we can explicitly write down the sum and then take the continuum limit.

The first thing to note is that the expectation value of  $(\Delta W)^2$  is equal to the variance of  $\Delta W$ , because  $\langle \Delta W \rangle = 0$ . Thus  $\langle (\Delta W)^2 \rangle = \Delta t$ . This tells us immediately that the expectation value of  $(\Delta W)^2$  does not vanish with respect to the time-step,  $\Delta t$ , and so the sum of these increments will not vanish when we sum over all the time-steps and take the infinitesimal limit. In fact, the expectation value of the sum

of all the increments  $(dW)^2$  from 0 to  $T$  is simply  $T$ :

$$\left\langle \int_0^T (dW)^2 \right\rangle = \int_0^T \langle (dW)^2 \rangle = \int_0^T dt = T. \quad (3.22)$$

Now let us see what the *variance* of the sum of all the  $(\Delta W)^2$  is. As we will see shortly, this drops to zero in the continuum limit, so that the integral of all the  $(dW)^2$  is not random at all, but deterministic! The reason for this is exactly the same reason that the variance of the average of  $N$  independent identical random variables goes to zero as  $N$  tends to infinity (see Section 1.5). If we sum  $N$  random variables together, but divide each by  $N$  so as to keep the mean fixed as  $N$  increases, then the variance of each variable drops as  $1/N^2$ . Because of this the sum of the  $N$  variances goes to zero as  $N \rightarrow \infty$ . This is exactly what happens to the sum of all the  $(\Delta W)^2$ , since the mean of each is equal to  $\Delta t = T/N$ . The sum of the means remains fixed at the total time interval  $T = N\Delta t$ , but the sum of the variances drops as  $1/N$ .

To explicitly calculate the variance of  $\int_0^T (dW)^2$ , we note that since  $\langle (\Delta W)^2 \rangle$  is proportional to  $\Delta t$ , the variance of  $(\Delta W)^2$  must be proportional to  $(\Delta t)^2$ . (We can in fact calculate it directly using the probability density for  $\Delta W$ , and the result is  $V[(\Delta W)^2] = 2(\Delta t)^2$ .) So, as discussed above, the variance of  $(\Delta W)^2$  is proportional to  $1/N^2$ :

$$V[(\Delta W)^2] = 2(\Delta t)^2 = 2 \frac{T^2}{N^2}. \quad (3.23)$$

The variance of the sum of all the  $(\Delta W)^2$  is thus

$$V \left[ \sum_{n=0}^{N-1} (\Delta W)^2 \right] = \sum_{n=0}^{N-1} V[(\Delta W)^2] = \sum_{n=0}^{N-1} 2(\Delta t)^2 = 2N \left( \frac{T}{N} \right)^2 = \frac{2T^2}{N}. \quad (3.24)$$

And hence the integral of all the  $(dW)^2$  is

$$\lim_{N \rightarrow \infty} V \left[ \sum_{n=0}^{N-1} (\Delta W)^2 \right] = \lim_{N \rightarrow \infty} \frac{2T^2}{N} = 0. \quad (3.25)$$

Since the integral of all the  $(dW)^2$  is deterministic, it is equal to its mean,  $T$ . That is,

$$\int_0^T (dW)^2 = T = \int_0^T dt. \quad (3.26)$$

Thus we have the surprising result that  $dW^2 = dt$ , a result officially known as **Ito's lemma**. We will refer to it here as **Ito's rule**. It is the fundamental rule for solving stochastic differential equations that contain Gaussian noise.

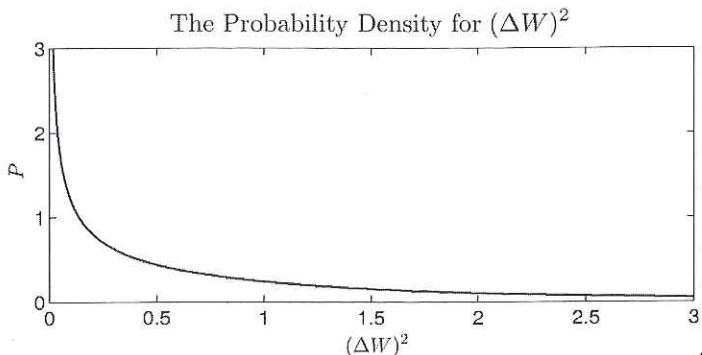


Figure 3.1. The probability density for  $(\Delta W)^2$ . The mean is  $\langle (\Delta W)^2 \rangle = \Delta t$ .

If you are not convinced that  $(dW)^2 = dt$  merely by calculating the variance of the integral of  $(dW)^2$ , as we have done above, we can do much better than this. It is not difficult to calculate the entire probability density for the integral of all the  $(dW)^2$ , and this shows us that the integral is exactly equal to the time interval  $T$ . To do this we use the characteristic function. Let us denote the  $n$ th square increment,  $(\Delta W_n)^2$ , as  $Z_n$ , and the sum of the square increments as the random variable  $Y(T)$ :

$$Y(T) = \sum_{n=0}^{N-1} Z_n = \sum_{n=0}^{N-1} [\Delta W_n]^2. \quad (3.27)$$

To calculate the probability density of  $Y(T)$ , we need to know first the probability density for  $Z_n \equiv (\Delta W_n)^2$ . Using the probability density for  $\Delta W$  (Eq. (3.7)) and the method for applying a transformation to a random variable as described in Section 1.6, we obtain

$$P(Z_n) = \frac{e^{-Z_n/(2\Delta t)}}{\sqrt{8\pi \Delta t}} \cdot \mathcal{D}(\Delta W_n) \frac{\partial(\Delta W_n)}{\partial Z_n} \quad (3.28)$$

This probability density is shown in Figure 3.1. We now take the Fourier transform of this to get the characteristic function, which is

$$\chi_{Z_n}(s) = \frac{1}{\sqrt{1 - 2is\Delta t}} = \frac{1}{\sqrt{1 - 2isT/N}}. \quad (3.29)$$

The characteristic function for  $Y(T) = \sum_{n=1}^N Z_n$ , is then the product of the characteristic functions for each of the  $Z_n$ . Hence

$$\chi_Y(s) = \left[ \frac{1}{\sqrt{1 - 2isT/N}} \right]^N \cdot \underbrace{\int_0^\infty e^{-x^2/(2\Delta t)} e^{isx} dx}_{\sqrt{\pi\Delta t}} \cdot \underbrace{\int_0^\infty e^{-x^2/(2\Delta t)} e^{isx} dx}_{\sqrt{\pi\Delta t}} \cdots \underbrace{\int_0^\infty e^{-x^2/(2\Delta t)} e^{isx} dx}_{\sqrt{\pi\Delta t}} = \sqrt{N} \sqrt{1 - 2isT}^{-N}. \quad (3.30)$$

Now we can quite easily take the continuum limit, which is

$$\chi_Y(s) = \lim_{N \rightarrow \infty} \left[ \frac{1}{\sqrt{1 - 2isT/N}} \right]^N = \lim_{N \rightarrow \infty} \left[ 1 + \frac{(isT)}{N} \right]^N = e^{isT}. \quad (3.31)$$

In the first step we have used the binomial approximation for the square root,

$$\sqrt{1 - x} \approx 1 + x/2, \quad (3.32)$$

because  $N$  is large. In the second step we have used the definition of the exponential function,  $e^x \equiv \lim_{N \rightarrow \infty} (1 + x/N)^N$ . Finally, we can now obtain the probability density for  $Y(T)$  by taking the inverse Fourier transform. This gives

$$P(Y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isY} \chi(s) ds = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{is(T-Y)} ds = \delta(T - Y). \quad (3.33)$$

The probability density for  $Y$  at time  $T$  is thus a delta function centered at  $T$ . Recall from Chapter 1 that a delta function  $\delta(x - a)$  is an infinitely sharp function which is zero everywhere except at  $x = a$ . It is infinitely tall at  $x = a$  so that the area underneath it is unity:  $\int_{-\infty}^{\infty} \delta(x - a) dx = 1$ . The fact that the probability density for  $Y$  is a delta function centered at  $T$  means that  $Y$  is exactly  $T$  and has no randomness.

*Summary:* we have now seen that the result of summing the Wiener increments  $(\Delta W)^2$ , in the infinitesimal limit, over the time interval  $T$  is simply  $T$ . This is just what we get if we sum  $dt$  over the same time interval. Thus, we have the surprising result that, in the continuum limit,

$$(dW)^2 = dt. \quad (3.34)$$

This relation, which we will refer to as *Ito's rule*, is a departure from the usual rules of calculus. This means that whenever  $(dW)^2$  appears in the process of solving a stochastic differential equation (SDE), it cannot be discarded, as can terms that are second and higher order in  $dt$ . However, it turns out that all other terms, such as products of the form  $dt^n dW^m$ , do vanish in the infinitesimal limit. The only terms that ever contribute to the solution of an SDE are  $dt$ ,  $dW$  and  $(dW)^2$ . Thus Ito's rule is the only additional rule that we will need to know to manipulate SDEs. The calculus of stochastic differential equations (at least in the form we have derived it here) is called *Ito calculus*.

### 3.5 Ito's formula: changing variables in an SDE

In solving ordinary differential equations in Chapter 2 we were able to ignore terms that were higher than first order in  $dt$ . Now that we are working with stochastic differential equations we will have to keep all terms that are first order in  $dt$  and

$dW$ , and all terms that are second order in  $dW$ . In fact, wherever we find terms that are second order in  $dW$  we can simply replace them with  $dt$ . We need to do this any time we have an SDE for a variable  $x$ , and wish to know the resulting SDE for a variable that is some function of  $x$ . To see how this works, consider a simple example in which we want to know the differential equation for  $y = x^2$ .

We must first work out the relationship between the increment of  $y$ ,  $dy$  and the increment of  $x$ . We have

$$\begin{aligned} dy &\equiv y(t+dt) - y(t) = x(t+dt)^2 - x(t)^2 \\ &= (x+dx)^2 - x^2 \\ &= x^2 + 2xdx + (dx)^2 - x^2 \\ &= 2xdx + (dx)^2. \end{aligned} \quad (3.35)$$

We see from this example that when we have a nonlinear function of a stochastic variable  $x$ , the second power of  $dx$  appears in the increment for that function. If  $x$  were deterministic then  $(dx)^2$  would vanish in the continuum limit, and we would have the usual rule of calculus, being

$$dy = 2xdx \quad \text{or} \quad \frac{dy}{dx} = 2x. \quad (3.36)$$

However, when  $x$  obeys the stochastic differential equation

$$dx = f dt + g dW, \quad (3.37)$$

we have

$$\begin{aligned} dy &= 2xdx + (dx)^2 \\ &= 2x(f dt + g dW) + g^2(dW)^2 \\ &= (2fx + g^2)dt + 2xgdW. \end{aligned} \quad (3.38)$$

This is Ito's rule in action.

Fortunately there is a simple way to calculate the increment of any nonlinear function  $y(x)$  in terms of the first and second powers of the increment of  $x$ . All we have to do is use the Taylor series expansion for  $y(x)$ , truncated at the second term:

$$dy = \left( \frac{\partial f}{\partial x} \right) dx + \frac{1}{2} \left( \frac{\partial^2 f}{\partial x^2} \right) (dx)^2. \quad (3.39)$$

If  $y$  is also an explicit function of time as well as  $x$ , then this becomes

$$dy = \left( \frac{\partial f}{\partial x} \right) dx + \left( \frac{\partial f}{\partial t} \right) dt + \frac{1}{2} \left( \frac{\partial^2 f}{\partial x^2} \right) (dx)^2. \quad (3.40)$$

We will refer to Eqs. (3.39) and (3.40) as *Ito's formula*.

$$dy = f dx + g dW + \frac{g^2}{2} dW^2$$

### 3.6 Solving some stochastic equations

#### 3.6.1 The Ornstein–Uhlenbeck process

We now turn to the problem of obtaining analytic solutions to stochastic differential equations. It turns out that there are very few that can be solved analytically, and we will examine essentially all of these in this chapter. The first example is that of a linear differential equation driven by “additive noise”. This is

$$dx = -\gamma x dt + g dW. \quad (3.41)$$

The term “additive noise” refers to the fact that the noise does not itself depend on  $x$ , but is merely added to any other terms that appear in the equation for  $dx$ . This equation is called the *Ornstein–Uhlenbeck equation*, and its solution is called the Ornstein–Uhlenbeck process. To solve Eq. (3.41) we note first that we know the solution to the deterministic part of the equation  $dx = -\gamma x dt$ , being  $x(t) = x_0 e^{-\gamma t}$ . We now change variables in the SDE to  $y = x e^{\gamma t}$  in the hope that this will simplify the equation. We can use Ito's formula for changing variables, as given in Eq. (3.40), but you may also find it instructive to do it explicitly. The latter method gives

$$\begin{aligned} dy &= y(x(t+dt), t+dt) - y(t) \\ &= y(x+dx, t+dt) - y(t) \\ &= (x+dx)e^{\gamma(t+dt)} - x e^{\gamma t} \\ &= xe^{\gamma t}\gamma dt + e^{\gamma t}(1+\gamma dt)dx \\ &= xe^{\gamma t}\gamma dt + e^{\gamma t}dx \\ &= \gamma y dt + e^{\gamma t}dx. \end{aligned} \quad (3.42)$$

In this calculation we have used the fact that the product of any infinitesimal increment with  $dt$  is zero. Substituting this expression for  $dy$  into the SDE for  $x$  gives

$$\begin{aligned} dy &= \gamma y dt + e^{\gamma t}dx \\ &= \gamma y dt + e^{\gamma t}(-\gamma x dt + g dW) \\ &= g e^{\gamma t} dW. \end{aligned} \quad (3.43)$$

Now we have an equation that is easy to solve! To do so we merely sum all the stochastic increments  $dW$  over a finite time  $t$ , noting that each one is multiplied by  $g e^{\gamma t}$ . The result is

$$y(t) = y_0 + g \int_0^t e^{\gamma s} dW(s). \quad (3.44)$$

This may look rather strange at first. So what *is* this stochastic integral? Well, it is a sum of Gaussian random variables, so it is itself merely a Gaussian random variable. Thus all we need to do is calculate its mean and variance. To do so we return to discrete time-steps so that we can calculate everything explicitly as we have done before. Thus we have

$$y(t) = y(N\Delta t) = y_0 + \lim_{N \rightarrow \infty} g \sum_{n=0}^{N-1} e^{\gamma n \Delta t} \Delta W_n, \quad (3.45)$$

where we have defined  $\Delta t = t/N$ . The variance of  $Y(t)$  is simply the sum of the variances of each of the random variables given by  $Y_n = g e^{\gamma n \Delta t} \Delta W_n$ . Since multiplying a random variable by a number  $c$  changes its variance by a factor of  $c^2$ , the variance of  $Y_n$  is  $g^2 e^{2\gamma n \Delta t} \Delta t$ . So we have

$$\begin{aligned} V[y(t)] &= \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} g^2 e^{2\gamma n \Delta t} \Delta t, \\ &= g^2 \int_0^t e^{2\gamma s} ds = \frac{g^2}{\gamma} (e^{2\gamma t} - 1). \end{aligned} \quad (3.46)$$

Similar reasoning shows that the mean of the stochastic integral in Eq. (3.44) is zero, and so  $\langle y(t) \rangle = y_0$ . This completely determines  $y(t)$ .

To obtain the solution of our original differential equation (Eq. (3.41)), we transform back to  $x$  using  $x = ye^{-\gamma t}$ . The solution is

$$x(t) = x_0 e^{-\gamma t} + g e^{-\gamma t} \int_0^t e^{\gamma s} dW(s) = x_0 e^{-\gamma t} + g \int_0^t e^{\gamma(s-t)} dW(s). \quad (3.47)$$

It is worth noting that the solution we have obtained is exactly the same solution that we would get if we replaced the driving noise by a deterministic function of time,  $f(t)$ . This would mean replacing  $dW$  with  $f(t)dt$ . The solution has exactly the same form because the method we used above to solve the stochastic equation was exactly the same method that we used in Chapter 2 to solve the same equation but with deterministic driving. That is, to solve this stochastic equation we do not need to use Ito calculus – normal calculus is sufficient. This can be seen immediately by noting that nowhere did  $(dW)^2$  appear in the analysis, because  $d^2y/dx^2 = 0$ . This is always true if the term giving the driving noise in the equation (the term containing  $dW$ ) does not contain  $x$ . When the driving noise does depend on  $x$ , then we cannot obtain the solution by assuming that  $dW = f(t)dt$  for some deterministic function  $f(t)$ , and must use Ito calculus to get the solution. In the next section we show how to solve the simplest equation of this type.

We can also solve the Ornstein–Uhlenbeck stochastic equation when  $\gamma$  and  $g$  are functions of time, using essentially the same method. That is, we change to a

new variable that has the deterministic part of the dynamics removed, and proceed as before. We leave the details of the derivation as an exercise. The solution in this case is

$$x(t) = x_0 e^{-\Gamma(t)} + \int_0^t e^{\Gamma(s)-\Gamma(t)} g(s) dW(s), \quad (3.48)$$

where

$$\Gamma(t) = \int_0^t \gamma(s) ds. \quad (3.49)$$

### 3.6.2 The full linear stochastic equation

This is the equation

$$dx = -\gamma x dt + g x dW. \quad \text{Solve this using Ito} \quad (3.50)$$

To solve it we can do one of two things. We can change variables to  $y = \ln x$ , or we can use the more direct approach that we used in Section 1.2. Using the latter method, we rewrite the differential equation as an exponential:

$$x(t+dt) = x + dx = [1 - \gamma dt + gdW]x = e^{[-\gamma + g^2/2]dt + gdW}x, \quad (3.51)$$

where we have been careful to expand the exponential to second order in  $dW$ . We now apply this relation repeatedly to  $x(0)$  to obtain the solution. For clarity we choose a finite time-step  $\Delta t = t/N$ , and we have

$$\begin{aligned} x(t) &= \lim_{N \rightarrow \infty} \left( \prod_{n=1}^N e^{-(\gamma + g^2/2)\Delta t + g \Delta W_n} \right) x(0) \\ &= \lim_{N \rightarrow \infty} \exp \left\{ -(\gamma + g^2/2)N\Delta t + g \sum_n \Delta W_n \right\} x(0) \\ &= \exp \left\{ -(\gamma + g^2/2)t + g \int_0^t dW \right\} x(0) \\ &= e^{-(\gamma + g^2/2)t + g W(t)} x(0). \end{aligned} \quad (3.52)$$

The random variable  $x(t)$  is therefore the exponential of a Gaussian random variable.

We can use the same method to solve the linear stochastic equation when  $\gamma$  and  $g$  are functions of time, and we leave the details of this calculation as an exercise. The solution is

$$x(t) = e^{-\Gamma(t)-H(t)+Z(t)} x(0), \quad (3.53)$$

where

$$\Gamma(t) = \int_0^t \gamma(s)ds, \quad (3.54)$$

$$H(t) = (1/2) \int_0^t g^2(s)ds, \quad (3.55)$$

and  $Z(t)$  is the random variable

$$Z(t) = \int_0^t g(s)dW(s). \quad (3.56)$$

### 3.6.3 Ito stochastic integrals

To summarize the results of Section 3.6.1 above, the stochastic integral

$$I(t) = \int_0^t f(s)dW(s) = \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} f(n\Delta t)\Delta W_n, \quad (3.57)$$

where  $f(t)$  is a deterministic function, is a Gaussian random variable with mean zero and variance

$$V[I(t)] = \int_0^t f^2(s)ds. \quad (3.58)$$

We now pause to emphasize an important fact about the stochastic integral in Eq. (3.57). Each term in the discrete sum contains a Wiener increment for a time-step  $\Delta t$ , multiplied by a function evaluated at the *start* of this time-step. This is a direct result of the way we have defined a stochastic differential equation – in fact, solutions to stochastic differential equations, as we have defined them, always involve integrals in which the function in the integral (the integrand) is evaluated at the start of each time-step. This has not been too important so far, but will become important when the integrand is itself a sum of Wiener increments. For example, if  $f(t)$  were itself a stochastic process, then it would be a function of all the increments  $\Delta W_j$  up until time  $t$ . So the fact that  $f(t)$  is evaluated at the start of the interval means that, for every term in the sum above,  $\Delta W_n$  is *independent* of all the increments that contribute to  $f(n\Delta t)$ . Because of this we have

$$\langle f(n\Delta t)\Delta W_n \rangle = \langle f(n\Delta t) \rangle \langle \Delta W_n \rangle = 0, \quad (3.59)$$

and thus  $\langle I(t) \rangle = 0$ , even if  $f(t)$  is a stochastic process. This will be important when we consider multiple stochastic integrals in Section 3.8.3, and modeling multiplicative noise in real systems in Section 5.3.

There are, in fact, other ways to define SDEs so that the resulting stochastic integrals are not defined in terms of the values of the integrand at the start of

each interval. They could, for example, be defined in terms of the values of the integrand in the center of each integral, or even at the end of each interval. Because there is more than one way to define a stochastic integral, those in which the integrand is evaluated at the start of each interval are called *Ito stochastic integrals*, and the corresponding SDEs *Ito stochastic equations*. The other versions of stochastic equations are much harder to solve. The main alternative to the Ito stochastic integral is called the *Stratonovich integral*, and as we will see in Section 5.3, this is useful for a specific purpose. Nevertheless, we note now that the alternative ways of defining stochastic integrals, and thus of defining SDEs, all give the same overall class of solutions – the various kinds of SDEs can always be transformed into Ito SDEs.

### 3.7 Deriving equations for the means and variances

So far we have calculated the means and variances of a stochastic process,  $x$ , at some time  $t$ , by first solving the SDE for  $x$  to obtain the probability density  $P(x, t)$ . However, there is another method that is sometimes useful if we are only interested in the low moments. If we have the SDE

$$dx = f(x, t)dt + g(x, t)dW \quad (3.60)$$

then we can obtain the differential equation for the mean of  $x$  by taking averages on both sides. This gives

$$d\langle x \rangle = \langle dx \rangle = \langle f(x, t) \rangle dt \quad (3.61)$$

because  $\langle dW \rangle = 0$ . We can rewrite this as

$$\frac{d\langle x \rangle}{dt} = \langle f(x, t) \rangle dt. \quad (3.62)$$

As an example of using this trick consider the linear SDE

$$dx = -\gamma xdt + \beta xdW. \quad (3.63)$$

We immediately obtain the equation for the mean, being

$$\frac{d\langle x \rangle}{dt} = -\gamma \langle x \rangle. \quad (3.64)$$

Thus

$$\langle x(t) \rangle = e^{-\gamma t} \langle x(0) \rangle. \quad (3.65)$$

To get the equation for the second moment, we first need to calculate the equation for  $x^2$ . This is

$$d\langle x^2 \rangle = -2\gamma x^2 dt + 2\beta x^2 dW + \beta^2 x^2 dt. \quad (3.66)$$

By taking the mean on both sides we get the differential equation for  $\langle x^2 \rangle$ , which is

$$d\langle x^2 \rangle = -2(\gamma - \beta^2)\langle x^2 \rangle dt. \quad (3.67)$$

We can now obtain the differential equation for the variance, by first noting that

$$\frac{dV_x}{dt} = \frac{d\langle x^2 \rangle}{dt} + \frac{d\langle x \rangle^2}{dt} = \frac{d\langle x^2 \rangle}{dt} + 2\langle x \rangle \frac{d\langle x \rangle}{dt}. \quad (3.68)$$

So using the differential equations that we have derived for the mean and second moment, that for the variance is

$$\frac{dV_x}{dt} = -2(\gamma - \beta^2)\langle x^2 \rangle - 2\gamma\langle x \rangle^2 = -2(\gamma - \beta^2)V_x - 2(2\gamma - \beta^2)\langle x \rangle^2. \quad (3.69)$$

Since we know  $\langle x(t) \rangle$  we can solve this using the techniques in Chapter 2.

In general, for a nonlinear stochastic equation, the equation of motion for the mean will contain the second moment(s), and those for the second moments will include the third moments, and so. The result is an infinite hierarchy of equations, which cannot be solved exactly.

### 3.8 Multiple variables and multiple noise sources

#### 3.8.1 Stochastic equations with multiple noise sources

Stochastic equations can, of course, be driven by more than one Gaussian noise source. These noise sources can be independent, or mutually correlated, but as we now explain, all correlated Gaussian noise sources can be obtained in a simple way from independent sources. We can solve stochastic differential equations driven by two (or more) independent Wiener noises using the same methods as those described above for a single noise source, along with the additional rule that the product of the increments of different noise sources are zero. As an example, to solve the equation

$$dx = f(x, t)dt + g_1(x, t)dW_1 + g_2(x, t)dW_2, \quad (3.70)$$

the Ito rules are

$$(dW_1)^2 = (dW_2)^2 = dt, \quad (3.71)$$

$$dW_1 dW_2 = 0. \quad (3.72)$$

To obtain two Wiener noise processes that are correlated, all we need to do is to form linear combinations of independent Wiener processes. If we define noise

sources  $dV_1$  and  $dV_2$  by

$$\begin{pmatrix} dV_1 \\ dV_2 \end{pmatrix} = M \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix} = \begin{pmatrix} \sqrt{1-\eta^2} & \eta \\ \eta & \sqrt{1-\eta^2} \end{pmatrix} \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix}, \quad (3.73)$$

with  $-1 \leq \eta \leq 1$ , then  $dV_1$  and  $dV_2$  are correlated even though  $dW_1$  and  $dW_2$  are not. The covariance matrix for  $dW_1$  and  $dW_2$  is  $I dt$  (where  $I$  is the two-by-two identity matrix), and that for  $dV_1$  and  $dV_2$  is

$$\begin{aligned} \begin{pmatrix} \langle (dV_1)^2 \rangle & \langle dV_1 dV_2 \rangle \\ \langle dV_1 dV_2 \rangle & \langle (dV_2)^2 \rangle \end{pmatrix} &= \left\langle \begin{pmatrix} dV_1 \\ dV_2 \end{pmatrix} (dV_1, dV_2) \right\rangle \\ &= \left\langle \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix} MM^T (dW_1, dW_2) \right\rangle \\ &= \begin{pmatrix} 1 & C \\ C & 1 \end{pmatrix} dt, \end{aligned} \quad (3.74)$$

where

$$C = \eta\sqrt{1-\eta^2}. \quad (3.75)$$

Since the means of  $dV_1$  and  $dV_2$  are zero, and because we have pulled out the factor of  $dt$  in Eq. (3.74),  $C$  is in fact the correlation coefficient of  $dV_1$  and  $dV_2$ .

The set of Ito calculus relations for  $dV_1$  and  $dV_2$  are given by essentially the same calculation:

$$\begin{pmatrix} (dV_1)^2 & dV_1 dV_2 \\ dV_1 dV_2 & (dV_2)^2 \end{pmatrix} = \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix} MM^T (dW_1, dW_2) \\ = \begin{pmatrix} 1 & C \\ C & 1 \end{pmatrix} dt. \quad (3.76)$$

If we have a stochastic equation driven by two correlated noise sources, such as

$$dx = f(x, t)dt + g_1(x, t)dV_1 + g_2(x, t)dV_2 = f(x, t)dt + \mathbf{g} \cdot \mathbf{dV}, \quad (3.77)$$

with  $\mathbf{dV} \equiv (dV_1, dV_2)^T$ , then we can always rewrite this in terms of the uncorrelated noises:

$$dx = f(x, t)dt + \mathbf{g} \cdot \mathbf{dV} = f(x, t)dt + \mathbf{g}^T \mathbf{M} \mathbf{dW}, \quad (3.78)$$

where we have defined  $\mathbf{dW} = (dW_1, dW_2)^T$ .

More generally, we can always write  $N$  correlated Gaussian noise processes,  $\mathbf{dV}$ , in terms of  $N$  independent Wiener processes,  $\mathbf{dW}$ . If we want the processes  $\mathbf{dV}$  to have the covariance matrix  $C$ , so that

$$\mathbf{dV} \mathbf{dV}^T = C dt, \quad (3.79)$$

then we define

$$d\mathbf{V} = M d\mathbf{W}, \quad (3.80)$$

where  $M$  is the square root of  $C$ . One can calculate the symmetric matrix  $M$  from  $C$  by using the definition of a function of a matrix given in Chapter 2: one diagonalizes  $C$  and then takes the square root of all the eigenvalues to construct  $M$ . (Note. Actually  $M$  does not have to be symmetric. If  $M$  is not symmetric, then  $C = MM^T$ .)

To summarize, when considering stochastic equations driven by multiple Gaussian noise sources, we only ever need to consider equations driven by independent Wiener processes.

### 3.8.2 Ito's formula for multiple variables

A general Ito stochastic differential equation that has multiple variables can be written in the vector form

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + G(\mathbf{x}, t)d\mathbf{W}. \quad (3.81)$$

Here  $\mathbf{x}$ ,  $\mathbf{f}$ , and  $d\mathbf{W}$  are the vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \mathbf{f} = \begin{pmatrix} f_1(\mathbf{x}, t) \\ f_2(\mathbf{x}, t) \\ \vdots \\ f_N(\mathbf{x}, t) \end{pmatrix} \quad d\mathbf{W} = \begin{pmatrix} dW_1 \\ dW_2 \\ \vdots \\ dW_M \end{pmatrix}, \quad (3.82)$$

where the  $dW_i$  are a set of mutually independent noise sources. They satisfy

$$dW_i dW_j = \delta_{ij} dt. \quad (3.83)$$

The symbol  $G$  is the  $N \times M$  matrix

$$G(\mathbf{x}, t) = \begin{pmatrix} G_{11}(\mathbf{x}, t) & G_{12}(\mathbf{x}, t) & \dots & G_{1M}(\mathbf{x}, t) \\ G_{21}(\mathbf{x}, t) & G_{22}(\mathbf{x}, t) & \dots & G_{2M}(\mathbf{x}, t) \\ \vdots & \vdots & \ddots & \\ G_{N1}(\mathbf{x}, t) & G_{N2}(\mathbf{x}, t) & \dots & G_{NM}(\mathbf{x}, t) \end{pmatrix}. \quad (3.84)$$

To determine Ito's formula for transforming stochastic equations involving multiple variables, all we have to do is use the multi-variable Taylor expansion. Let us say that we wish to transform from a set of variables  $\mathbf{x} = (x_1, \dots, x_N)^T$ , to a set of variables  $\mathbf{y} = (y_1, \dots, y_L)^T$ , where each of the  $y_i$  is a function of some or all of the

$x_i$ , and of time,  $t$ . Using the Taylor expansion we have

$$dy_i = \sum_{j=1}^N \frac{\partial y_i}{\partial x_j} dx_j + \frac{\partial y_i}{\partial t} dt + \frac{1}{2} \sum_{k=1}^N \sum_{j=1}^N \frac{\partial^2 y_i}{\partial x_k \partial x_j} dx_k dx_j, \quad i = 1, \dots, L. \quad (3.85)$$

This is all one requires to transform variables for any SDE. If we substitute the multivariate SDE given by Eq. (3.81) into the Taylor expansion, and use the Ito rules given in Eq. (3.83) we obtain

$$\begin{aligned} dy_i &= \sum_{j=1}^N \frac{\partial y_i}{\partial x_j} \left( f_j(\mathbf{x}, t)dt + \sum_k G_{jk}(\mathbf{x}, t)dW_k \right) + \frac{\partial y_i}{\partial t} dt \\ &\quad + \frac{1}{2} \sum_{k=1}^N \sum_{j=1}^N \frac{\partial^2 y_i}{\partial x_k \partial x_j} \left( \sum_{m=1}^M G_{jm} G_{km} \right) dt, \quad i = 1, \dots, L. \end{aligned} \quad (3.86)$$

### 3.8.3 Multiple Ito stochastic integrals

As we have seen in various examples above, the integrals that appear in the solutions to Ito stochastic differential equations for a single variable have the form

$$\int_0^t f(s)dW(s) = \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} f(n\Delta t)\Delta W_n, \quad (3.87)$$

for some function  $f(t)$ , where  $\Delta t = t/N$ . In the above summation  $f(n\Delta t)$  is the value of the function  $f(t)$  at the *start* of the time interval to which the Wiener increment  $\Delta W_n$  corresponds. As we will see below, this fact becomes important when evaluating multiple stochastic integrals.

When solving Ito equations that have multiple variables, or multiple noise processes, the solutions are in general multiple integrals that involve one or more Wiener processes. It is therefore useful to know how to handle such integrals. For example, we might be faced with the integral

$$I = \int_0^T \int_0^t f(s)dW(s)dt. \quad (3.88)$$

At first sight this may look confusing, but it is easy to evaluate. By discretizing the integral (that is, writing it as a double summation) it becomes clear that we can exchange the order of integration, just as we can for a regular integral. Thus

$$I = \int_0^T \left[ \int_0^t f(s)ds \right] dW(t). \quad (3.89)$$

Since this is merely a function of  $t$  integrated over the Wiener process, we already know from Section 3.6.1 that it is a Gaussian random variable with mean zero and variance

$$V[I] = \int_0^T \left[ \int_0^t f(s) dW(s) \right]^2 dt. \quad (3.90)$$

Now, what about double Ito integrals in which both integrals are over Wiener processes? In this case we have two possibilities: the two integrals may involve the *same* Wiener process, or two independent Wiener processes. Note that in both cases we have a double sum of *products* of Wiener increments, so the resulting random variable is no longer Gaussian. In general there is no analytic solution to integrals of this form, but it is not difficult to calculate their expectation values. Before we show how to do this, we note that there is a special case in which a double integral of a single Wiener process can be evaluated exactly. This special case is

$$I = \int_0^T \left[ \int_0^t f(s) dW(s) \right] f(t) dW(t). \quad (3.91)$$

To solve this we first define the random variable  $Z(t)$  as the value of the inner integral:

$$Z(t) = \int_0^t f(s) dW(s). \quad (3.92)$$

Note that  $Z(t)$  is Gaussian with zero mean and variance  $V = \int_0^t f^2(s) ds$ . We now discretize the double integral:

$$I = \int_0^T Z(t) dW(t) = \lim_{\Delta t \rightarrow 0} \sum_{n=0}^{N-1} Z_n f_n \Delta W_n, \quad (3.93)$$

where  $f_n \equiv f(n\Delta t)$ , and rewrite the sum as follows:

$$\begin{aligned} I &= \lim_{\Delta t \rightarrow 0} \sum_{n=0}^{N-1} (Z_n + f_n \Delta W_n)^2 - Z_n^2 - (f_{n-1} \Delta W_n)^2 \\ &= \lim_{\Delta t \rightarrow 0} \sum_{n=0}^{N-1} \Delta(Z_n^2) - f_n^2 (\Delta W_n)^2 = \int_0^T d(Z_n^2(t)) - \int_0^T f^2(t) (dW)^2 \\ &= Z^2(T) - \int_0^T f^2(t) dt. \end{aligned} \quad (3.94)$$

This is an exact expression for  $I$ , since we can easily obtain an exact expression for the probability density of  $Z^2$ .

While we cannot obtain an analytic expression for the integral

$$J = \int_0^T \left[ \int_0^T f(s) dW(s) \right] g(t) dW(t) \quad (3.95)$$

(where once again both integrals contain increments of the same Wiener process), we can calculate the expectation value of  $J$ . Note that in the definition of  $J$  we have fixed the upper limits of both the inner and outer integrals to be  $T$ . We will consider replacing the upper limit of the inner integral with  $t$  shortly. Discretizing  $J$ , and taking the average, we have

$$\begin{aligned} \langle J \rangle &= \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} \sum_{m=1}^N f_m g_n \langle \Delta W_m \Delta W_n \rangle \\ &= \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} f_n g_n \langle (\Delta W_n)^2 \rangle = \int_0^T f(t) g(t) dt. \end{aligned} \quad (3.96)$$

We get this result because  $\langle \Delta W_m \Delta W_n \rangle$  is  $\Delta t$  if  $m = n$ , and zero otherwise.

Now let us see what happens if the upper limit of the inner integral is equal to the variable in the outer integral. In this case, the inner integral, being  $Z(t) = \int_0^t f(s) dW(s)$ , only contains increments of the Wiener process up until time  $t$ , and the Wiener increment in the outer integral that multiplies  $Z(t)$  is the increment just *after* time  $t$ . (It is the increment from  $t$  to  $t + dt$ .) Discretizing the integral we have

$$\langle I \rangle = \lim_{N \rightarrow \infty} \sum_{n=1}^N \left[ \sum_{m=1}^{n-1} f_{m-1} g_{n-1} \langle \Delta W_m \Delta W_n \rangle \right]. \quad (3.97)$$

Thus there are no terms in the double sum that have a product of Wiener increments in the *same* time interval, only those in different time intervals. As a result the expectation value is zero:

$$\langle I \rangle = \left\langle \int_0^T \left[ \int_0^t f(s) dW(s) \right] g(t) dW(t) \right\rangle = 0. \quad (3.98)$$

We can encapsulate the above results, Eq. (3.96) and Eq. (3.98), with the rule

$$\langle dW(s) dW(t) \rangle \equiv \delta(t-s) ds dt, \quad (3.99)$$

but this works *only* with the following additional rule: when one of the arguments in the  $\delta$ -function is equal to the upper or lower limit of the integral, we have

$$\int_L^U \delta(t-L) dt = f(L), \quad (3.100)$$

$$\int_L^U \delta(t-U) dt = 0. \quad (3.101)$$

This means that, for Ito integrals, we count all the area, or “weight”, of the  $\delta$ -function in Eq. (3.99) at the lower limit of an integral, and none at the upper limit.

### 3.8.4 The multivariate linear equation with additive noise

A multivariate linear system with additive noise is also referred to as a multivariate Ornstein–Uhlenbeck process. When there is only a single noise source this is described by the vector stochastic equation

$$d\mathbf{x} = F\mathbf{x}dt + \mathbf{g}(t)dW, \quad (3.102)$$

where  $\mathbf{x}$  and  $\mathbf{g}$  are vectors, and  $F$  is a matrix. In addition,  $F$  is a constant, and  $\mathbf{g}(t)$  is a function of  $t$  but not of  $\mathbf{x}$ . To solve this we can use exactly the same technique as that used above for the single variable Ornstein–Uhlenbeck process, employing what we know about solving linear vector equations from Chapter 2. The solution is

$$\mathbf{x}(t) = e^{Ft}\mathbf{x}(0) + \int_0^t e^{F(t-s)}\mathbf{g}(s)dW(s). \quad (3.103)$$

We can just as easily solve the multivariate Ornstein–Uhlenbeck process when there are multiple independent noises driving the system. In this case the stochastic equation is

$$d\mathbf{x} = F\mathbf{x}dt + G(t)\mathbf{dW}, \quad (3.104)$$

where  $\mathbf{dW} = (dW_1, dW_2, \dots, dW_N)$  is a vector of mutually independent Wiener noises  $dW_i$ . The solution is obtained in exactly the same way as before, and is

$$\mathbf{x}(t) = e^{Ft}\mathbf{x}(0) + \int_0^t e^{F(t-s)}G(s)\mathbf{dW}(s). \quad (3.105)$$

### 3.8.5 The full multivariate linear stochastic equation

The multivariate linear stochastic for a single noise source is

$$d\mathbf{x} = F\mathbf{x}dt + G\mathbf{x}dW, \quad (3.106)$$

and we will take the matrices  $F$  and  $G$  to be constant. It is not always possible to obtain an analytic solution to this equation. Whether or not an analytic solution exists depends on the *commutator*, or *commutation relations*, between  $F$  and  $G$ . Many readers will not be familiar with the term “commutation relations”, so we now explain what this means.

First, the *commutator* of  $F$  and  $G$  is defined as

$$[F, G] \equiv FG - GF. \quad (3.107)$$

If  $[F, G] = 0$  then  $FG = GF$ , and  $F$  and  $G$  are said to *commute*. Two arbitrary matrices  $F$  and  $G$  in general do not commute. The reason that the commutator of  $F$  and  $G$  is important, is because, when  $F$  and  $G$  do not commute, it is no longer true that  $e^F e^G = e^{F+G}$ . Recall that when we solved the linear stochastic equation for a single variable, we used the fact that  $e^a e^b = e^{a+b}$ , which is always true when  $a$  and  $b$  are numbers. This is still true for matrices when they commute, and as a result the solution to Eq. (3.106) when  $F$  and  $G$  commute is

$$x(t) = e^{(F+G^2/2)t+GW(t)}x(0), \quad [F, G] = 0. \quad (3.108)$$

To solve the vector linear stochastic equation when  $F$  and  $G$  do not commute, we need to know how  $e^F e^G$  is related to  $e^{F+G}$ . This relationship is called the Baker–Campbell–Hausdorff (BCH) formula, and is

$$e^F e^G = e^{F+G+Z_2+Z_3+Z_4+\dots+Z_\infty}, \quad (3.109)$$

where the additional terms in the sum are “repeated” commutators of  $F$  and  $G$ . The first two of these terms are

$$Z_2 = (1/2)[F, G], \quad (3.110)$$

$$Z_3 = (1/12)[F, [F, G]] + (1/12)[[F, G], G], \quad (3.111)$$

and higher terms become increasingly more complex [9, 10]. Because of this we cannot obtain a closed-form solution to Eq. (3.106) for every choice of the matrices  $F$  and  $G$ . However, if the commutator, or some repeated commutator, of  $F$  and  $G$  is sufficiently simple, then the infinite series in the BCH formula terminates at some point so that it has only a finite number of terms.

We illustrate how to solve Eq. (3.106) for the simplest example in which  $F$  and  $G$  do not commute. This is the case in which  $[F, G]$  commutes with both  $F$  and  $G$ . That is

$$[F, [F, G]] = [G, [F, G]] = 0. \quad (3.112)$$

In this case  $Z_3$  is zero. Because all higher-order terms involve the commutators of  $F$  and  $G$  with the terms in  $Z_3$ , these vanish too, and the only additional term that remains is  $Z_2 = [F, G]/2$ . The relationship between  $e^F e^G$  and  $e^{F+G}$  becomes

$$e^F e^G = e^{F+G+[F,G]/2} = e^{[F,G]/2} e^{F+G}. \quad (3.113)$$

The second relation is true because  $F + G$  commutes with  $[F, G]$ . Also, since  $[F, G] = -[G, F]$  we have

$$e^F e^G = e^{F+G-[F,G]/2} = e^{-[F,G]/2} e^{F+G}, \quad (3.114)$$

and putting the above relations together gives

$$e^F e^G = e^{[F,G]} e^G e^F. \quad (3.115)$$

To solve Eq. (3.106) we first write it as

$$\mathbf{x}(t + dt) = e^{\tilde{F}dt + GdW} \mathbf{x}(t) = e^{CdWdt} e^{\tilde{F}dt} e^{GdW} \mathbf{x}(t), \quad (3.116)$$

where we have defined  $C \equiv [F, G]$  and  $\tilde{F} \equiv F - G^2/2$  to simplify the notation. Here we have used  $(dW)^2 = dt$  and the above relations for the matrix exponential. We can now write the solution as

$$\mathbf{x}(t) = \lim_{N \rightarrow \infty} \prod_{n=1}^N \left( e^{C\Delta W_n \Delta t} e^{\tilde{F}\Delta t} e^{G\Delta W_n} \right) \mathbf{x}(0), \quad (3.117)$$

where as usual  $\Delta t = t/N$ . Now we must turn the products of exponentials into exponentials of sums, just like we did to solve the single variable equation. Let us see what to do when we have only two time-steps. In this case we have

$$\mathbf{x}(2\Delta t) = e^{C(\Delta W_1 + \Delta W_2)\Delta t} e^{\tilde{F}\Delta t} e^{G\Delta W_2} e^{\tilde{F}\Delta t} e^{G\Delta W_1} \mathbf{x}(0). \quad (3.118)$$

Note that because  $C$  commutes with all the other operators, we can always place all the terms containing  $C$  to the left, and combine them in a single exponential. We need to swap the middle two terms,  $e^{G\Delta W_2}$  and  $e^{\tilde{F}\Delta t}$ . If we do this then we have all the terms containing  $\tilde{F}$  to the left, which allows us to combine them, and all those containing the  $\Delta W$ s to the right, also allowing us to combine them. When we perform the swap, we get an additional term  $e^{-C\Delta t \Delta W}$  because  $[G, \tilde{F}] = [G, F] = -C$ . The result is

$$\begin{aligned} \mathbf{x}(2\Delta t) &= e^{C(\Delta W_1 + 2\Delta W_2)\Delta t} e^{-C\Delta W_2 \Delta t} e^{\tilde{F}\Delta t} e^{\tilde{F}\Delta t} e^{G\Delta W_2} e^{G\Delta W_1} \mathbf{x}(0) \\ &= e^{C(\Delta W_1 + \Delta W_2)} e^{-C\Delta W_2 \Delta t} e^{\tilde{F}2\Delta t} e^{G(\Delta W_1 + \Delta W_2)} \mathbf{x}(0). \end{aligned} \quad (3.119)$$

To obtain  $x(3\Delta t)$  we multiply this on the left by  $e^{C\Delta W_3 \Delta t} e^{\tilde{F}\Delta t} e^{G\Delta W_3}$ . Once again we must swap the two middle terms, bringing the term  $e^{\tilde{F}\Delta t}$  to the left. Doing this, which we leave as an exercise, the pattern becomes clear. We can now perform the swap operation  $N$  times, and the result is

$$\mathbf{x}(t) = \lim_{N \rightarrow \infty} \left( e^{C\Delta t \sum_n \Delta W_n} e^{-C\Delta t \sum_n n \Delta W_n} e^{\tilde{F}t} e^{G\Delta \sum_n W_n} \right) \mathbf{x}(0). \quad (3.120)$$

We can now take the limit as  $N \rightarrow \infty$ . The limits of the various sums are

$$\lim_{N \rightarrow \infty} \Delta t \sum_n \Delta W_n = \left( \lim_{\Delta t \rightarrow 0} \Delta t \right) \int_0^t dW(s) = 0, \quad (3.121)$$

$$\lim_{N \rightarrow \infty} \Delta t \sum_n n \Delta W_n = \lim_{N \rightarrow \infty} \sum_n (n \Delta t) (\Delta W_n) = \int_0^t s dW(s), \quad (3.122)$$

$$\lim_{N \rightarrow \infty} \sum_n \Delta W_n = \int_0^t dW(s) = W(t). \quad (3.123)$$

The solution to Eq. (3.106), when  $[F, G]$  commutes with  $F$  and  $G$ , is thus

$$\mathbf{x}(t) = e^{(F-G^2/2)t} e^{GW(t)} e^{-[F,G]Z(t)} \mathbf{x}(0), \quad (3.124)$$

where we have defined  $Z(t) = \int_0^t s dW(s)$ . The mean of  $Z(t)$  is zero, and the variance is  $t^3/3$ .

Note that in Section 3.4 above we said that we could always set the product  $dWdt$  to zero. This is true if this product appears in a stochastic differential equation, and this is usually all that matters. However, we see that in implementing the method we have described above for deriving an analytic solution to an SDE we could not drop terms proportional to  $dWdt$ : we had to keep these terms because during the repeated swap operations enough of them were generated to produce a non-zero result.

### 3.9 Non-anticipating functions

Before we finish with this chapter, it is worth defining the term *non-anticipating function*, also known as an *adapted process*, since it is used from time to time in the literature. Because a stochastic process,  $x(t)$ , changes randomly in each time-step  $dt$ , its value at some future time  $T$  is not known at the initial time. Once the future time has been reached, then all the values of the stochastic increments up until that time have been chosen, and the value of  $x(T)$  is known. So at the initial time our state of knowledge of  $x(T)$  is merely the probability density for  $x(T)$ , but at time  $T$ ,  $x(T)$  has a definite value that has been picked from this density. To sum this up, if the current time is  $t$ , then one knows all the random increments,  $dW$ , up until time time  $t$ , but none beyond that time.

We say that a function of time,  $f(t)$ , is *non-anticipating*, or *adapted to the process*  $dW$ , if the value of the function at time  $t$  depends only on the stochastic increments  $dW$  up until that time. Thus if  $x(t)$  is the solution to a stochastic equation drive by a stochastic process  $dW$ , then  $x(t)$  is a non-anticipating function, and any function of  $f(x, t)$  is also. When solving stochastic equations, one always deals with non-anticipating functions.

We can go further in this direction. At the initial time,  $t = 0$ , we have some probability density for  $x(T)$ , which we obtain by solving the SDE for  $x$ . However, at a time,  $t$ , where  $0 < t < T$ , then we know the increments up until time  $t$ , and thus the value of  $x(t)$ . While we still do not know the value of  $x(T)$ , our probability density for it will now be different – since we know some of the stochastic increments in the interval  $[0, T]$ , we have more information about the likely values of  $X(T)$ . Our probability density at time  $t$  for  $x(T)$  is therefore a *conditional* probability density, conditioned on a knowledge of the increments in the interval  $[0, t]$ . We will discuss more about how one calculates these conditional probability densities in the next chapter.

### Further reading

Our presentation of stochastic differential equations here is intentionally non-rigorous. Our purpose is to convey a clear understanding of SDEs and Ito calculus without getting bogged down by mathematical rigor. We discuss the concepts employed in rigorous treatments in Chapter 10. A rigorous mathematical account of Ito's rule and stochastic differential equations driven by Wiener noise may be found in, for example, *Stochastic Differential Equations: An Introduction with Applications* by Bernt Øksendal [11], *Brownian Motion and Stochastic Calculus* by Karatzas and Shreve [12], and the two-volume set *Diffusions, Markov Processes, and Martingales* by Rogers and Williams [13]. The first of these focusses more on modeling and applications, while the second two contain more theorems and details of interest to mathematicians. The techniques for obtaining analytic solutions to stochastic differential equations presented in this chapter are essentially exhaustive, as far as the author is aware. One further example of a solution to a vector linear stochastic equation containing more complex commutation relations than the one solved in Section 3.8.5 is given in [14], where the method presented in that section was introduced.

### Exercises

1. By expanding the exponential to second order, show that  $e^{\alpha dt - \beta^2 dt/2 + \beta dW} = 1 + \alpha dt + \beta dW$  to first order in  $dt$ .
2. If the discrete differential equation for  $\Delta x$  is

$$\Delta x = cx\Delta t + b\Delta W \quad (3.125)$$

and  $x(0) = 0$ , calculate the probability density for  $x(2\Delta t)$ .

3. If  $z(t) = \int_0^t t'^2 dW(t')$ , and  $x = az + t$ , what is the probability density for  $x$ ?

4.  $x(t) = e^{-b(W(t))^2}$ .
  - (i) What values can  $x$  take?
  - (ii) What is the probability density for  $x$ ?
  - (iii) Calculate  $\langle x^2 \rangle$  using the probability density for  $W$ .
  - (iv) Calculate  $\langle x^2 \rangle$  using the probability density for  $x$ .

5. Calculate the expectation value of  $e^{\alpha t + \beta W(t)}$ .
6. The random variables  $x$  and  $y$  are given by

$$x(t) = a + [W(t)]^2,$$

$$y(t) = e^{bW(t)}.$$

- (i) Calculate  $\langle x(t) \rangle$ .
- (ii) Calculate  $\langle y(t) \rangle$ .
- (iii) Calculate  $\langle x(t)y(t) \rangle$ .
- (iv) Are  $x(t)$  and  $y(t)$  independent?

7. The stochastic process  $x$  satisfies the stochastic equation

$$dx = -\gamma xdt + gdW. \quad (3.126)$$

- (i) Calculate the differential equations for  $\langle x \rangle$  and  $\langle x^2 \rangle$ .
- (ii) Calculate the differential equation for  $V(x) = \langle x^2 \rangle - \langle x \rangle^2$ , and solve it.

8. The stochastic differential equation for  $x$  is

$$dx = 3a(x^{1/3} - x)dt + 3\sqrt{ax^{2/3}}dW.$$

- (i) Transform variables to  $y = x^{1/3}$ , so as to get the stochastic equation for  $y$ .
- (ii) Solve the equation for  $y$ .
- (iii) Use the solution for  $y(t)$  to get  $x(t)$ , making sure that you write  $x(t)$  in terms of  $x_0 \equiv x(0)$ .

9. Solve the stochastic differential equation

$$dx = -\alpha t^2 xdt + gdW, \quad (3.127)$$

and calculate  $\langle x(t) \rangle$  and  $V[x(t)]$ . Hint. To do this, you use essentially the same method as for the Ornstein–Uhlenbeck equation: first solve the differential equation with  $g = 0$ , and this tells you how to make a transformation to a new variable, so that the new variable is constant when  $g = 0$ . The stochastic equation for the new variable is then easy to solve.

10. Linear stochastic equation with time-dependent coefficients: obtain the solution to the equation

$$dx = -f(t)xdt + g(t)xdW. \quad (3.128)$$

11. Solve the linear stochastic equation

$$dx = -\gamma x dt + g x dW, \quad (3.129)$$

by first changing variables to  $y = \ln x$ . Then calculate the probability density for the solution  $x(t)$ .

12. Solve the stochastic equations

$$dx = p dW \quad (3.130)$$

$$dp = -\gamma p dt. \quad (3.131)$$

13. Solve the stochastic equations

$$dx = pdt + \beta dV \quad (3.132)$$

$$dp = -\gamma dW, \quad (3.133)$$

where  $dV$  and  $dW$  are two mutually independent Wiener processes. That is,  $(dV)^2 = (dW)^2 = dt$  and  $dVdW = 0$ .

14. Solve the stochastic equation

$$dx = Axdt + BxdW, \quad (3.134)$$

where

$$A = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}, \quad B = \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix}. \quad (3.135)$$

## Further properties of stochastic processes

We have seen in the previous chapter how to define a stochastic process using a sequence of Gaussian infinitesimal increments  $dW$ , and how to obtain new stochastic processes as the solutions to stochastic differential equations driven by this Gaussian noise. We have seen that a stochastic process is a random variable  $x(t)$  at each time  $t$ , and we have calculated its probability density,  $P(x, t)$ , average  $\langle x(t) \rangle$  and variance  $V[x(t)]$ . In this chapter we will discuss and calculate some further properties of a stochastic process, in particular its *sample paths*, *two-time correlation function*, and power spectral density (or *power spectrum*). We also discuss the fact that Wiener noise is *white noise*.

### 4.1 Sample paths

A *sample path* of the Wiener process is a particular choice (or *realization*) of each of the increments  $dW$ . Since each increment is infinitesimal, we cannot plot a sample path with infinite accuracy, but must choose some time discretization  $\Delta t$ , and plot  $W(t)$  at the points  $n\Delta t$ . Note that in doing so, even though we do not calculate  $W(t)$  for the points in-between the values  $n\Delta t$ , the points we plot do lie precisely on a valid sample path, because we know precisely the probability density for each increment  $\Delta W$  on the intervals  $\Delta t$ . If we chose  $\Delta t$  small enough, then we cannot tell by eye that the resolution is limited. In Figure 4.1 we plot a sample path of the Wiener process.

If we want to plot a sample path of a more complex stochastic process, defined by an SDE that we cannot solve analytically, then we must use a numerical method to solve the SDE. In this case our solution at each time-step will be approximate, and thus the points we plot on the sample path will also be approximate. We discuss how to solve SDEs numerically in Chapter 6.

So far we have mainly thought of a stochastic process,  $x(t)$ , as a random variable at each time  $t$ . However, it is more accurate to think of it as being described by a

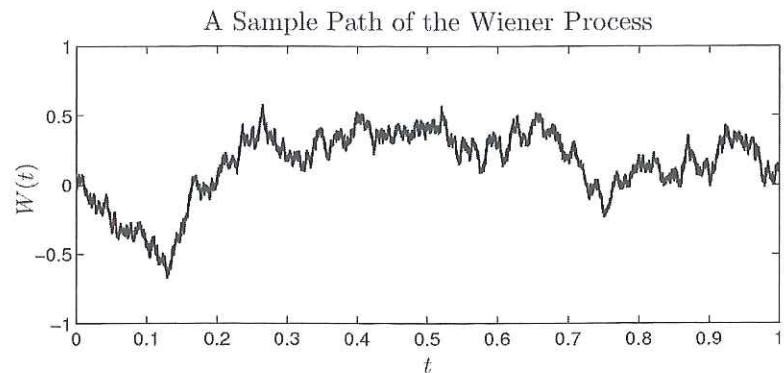


Figure 4.1. A sample path of the Wiener process.

collection of many different possible sample paths. Thus  $x(t)$  is actually a random *function*, whose possible values are the different samples paths. If we are only interested in the average value of  $x(t)$ , then we only need the probability density for  $x$  at each time  $t$ . If we want to calculate  $\langle x(t)x(t + \tau) \rangle$ , being the expectation value of the product of  $x$  with itself at a later time (which we will do in Section 4.4 below), then we need the joint probability density for  $x$  at the two times  $t$  and  $t + \tau$ . How much we need to know about the full probability density for the function  $x(t)$  over all the sample paths therefore depends on what we want to calculate.

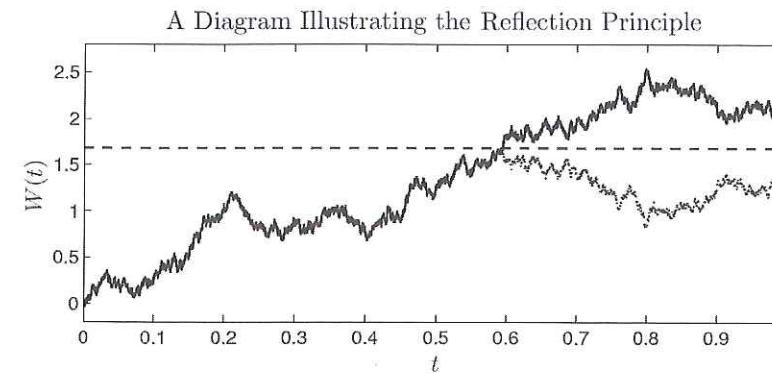
The sample paths of the Wiener process fluctuate on every time-scale: no matter how small an interval,  $\Delta t$ , we use, because successive Wiener increments  $\Delta W$  are independent,  $W(t)$  will change randomly from one time-step to the next. The result is that a sample path of the Wiener process is *not differentiable*. We can see this if we attempt to calculate the derivative:

$$\frac{dW}{dt} = \lim_{h \rightarrow 0} \frac{W(t+h) - W(t)}{h} = \lim_{h \rightarrow 0} \frac{W(h)}{h}. \quad (4.1)$$

But we know that  $W(h)$  is of order  $\sqrt{h}$  (because the standard deviation of  $W(h)$  is  $\sqrt{h}$ ), so we have

$$\frac{dW}{dt} = \lim_{h \rightarrow 0} \frac{W(h)}{h} \sim \lim_{h \rightarrow 0} \frac{1}{\sqrt{h}} \rightarrow \infty. \quad (4.2)$$

The fact that the Wiener process is not differentiable is the reason that it is possible to have the relation  $dW^2 = dt$  and thus break the usual rules of calculus. If the Wiener process had a derivative then all second-order terms would vanish, and we would not need to do all our calculations in terms of differentials. We could instead write our differential equations in terms of derivatives such as  $dx/dt = -\gamma x + g dW/dt$ , etc. (In fact, it turns out that there is a special case in which we can use the notation  $dW/dt$ , and this will be described in Section 4.6.)

Figure 4.2. Here we show a sample path of the Wiener process that crosses the value  $a$  at  $t = \tau$ . We also show a second sample path that is the same as the first one up until it crosses  $a$ , and then is the reflection of the first one about the horizontal line at  $a$  (dotted line).

## 4.2 The reflection principle and the first-passage time

Here we consider the time that it takes the Wiener process to reach a given value,  $a$ . (Recall that the Wiener process, by definition, starts with the value zero.) This is called the *first-passage time* to reach the “boundary”  $a$ , and we will denote it by  $T_a$ . Naturally, since the Wiener process is stochastic, each realization of the process will take a different time to reach the value  $a$ . So what we actually want to calculate is the probability density for  $T_a$ . This seems at first like a difficult thing to calculate, but it turns out that there is a clever trick (first realized by D. André in 1887) that makes this fairly simple (although the reasoning is a little involved). This trick is called the “reflection principle”. We note that our discussion of the reflection principle as a way of calculating first-passage times is included here mainly as a curiosity. In Section 7.7 we will discuss a more generally applicable, and possibly more straightforward way to calculate first-passage times, using Fokker–Planck equations.

Consider a sample path of the Wiener process that crosses the value  $a$  at some time  $\tau$ , as shown in Figure 4.2. We can obtain another sample path from this one by *reflecting* the first path about the horizontal line at height  $a$  after the sample path crosses  $a$ . This second sample path is the same as the first one up until the crossing time  $\tau$ , but is different from it after that time, and is shown in Figure 4.2 as a dotted line. Now, the important point about this construction is the following: since the Wiener process is *symmetric* (that is, it is just as likely to go up in any given time-step as to go down) the probability of the first sample path to occur is exactly the same as that of the second. This is the reflection principle. It states that if you reflect the sample path of a symmetric stochastic process from any point onwards, the new path has the same probability of occurring as the first path.

We will now use the reflection principle to calculate the probability density for the first-passage time. Consider a Wiener process that starts with the value 0 at time  $t = 0$ , and evolves for a time  $T$ . During this time it will reach some maximum value that will depend on the sample path that it follows. Call this maximum  $M_T$ . Let us now calculate the probability that this maximum  $M_T$  is greater than or equal to  $a$ . Examining the two paths shown in Figure 4.2, we see that in both cases the maximum value of the Wiener process is greater than or equal to  $a$ . However, in only one case (the first path, represented by the solid line) is the final value of the Wiener process greater than  $a$ . Thus for every sample path for which the final value of the Wiener process is greater than  $a$ , there are *two* paths for which  $M_T \geq a$ . Therefore we can conclude that

$$\text{Prob}(M_T \geq a) = 2 \times \text{Prob}(W(T) \geq a). \quad (4.3)$$

The probability that  $W(T)$  is greater than  $a$  is easier to calculate, because we know the probability density for  $W(T)$ . Thus

$$\text{Prob}(M_T \geq a) = 2 \times \text{Prob}(W(T) \geq a) = \frac{2}{\sqrt{2\pi T}} \int_a^\infty e^{-x^2/(2T)} dx. \quad (4.4)$$

This integral has no analytic solution, but can of course be evaluated numerically.

From the above result we can calculate the probability density for the first-passage time to reach  $a$ . We merely have to realize that the time it takes the Wiener process to reach  $a$  will be less than or equal to  $T$  if and only if the maximum of the Wiener process in time  $T$ ,  $M_T$  is greater than  $a$ . Hence we have

$$\text{Prob}(T_a \leq T) = \text{Prob}(M_T \geq a) = \frac{2}{\sqrt{2\pi T}} \int_a^\infty e^{-x^2/(2T)} dx. \quad (4.5)$$

Now the probability that  $T_a \leq T$  is simply the distribution function for  $T_a$ ,  $D(T)$ , and so the probability density for  $T_a$ ,  $P(T)$ , is

$$P(T) = \frac{dD(T)}{dT} = \frac{d}{dT} \left[ \sqrt{\frac{2}{\pi T}} \int_a^\infty e^{-x^2/(2T)} dx \right]. \quad (4.6)$$

To evaluate this derivative, we change variables under the integral to  $v = (Ta^2)/x^2$  (Note that  $a$  is positive, so the integration variable  $x$  is always positive, which ensures the mapping from  $x \rightarrow v$  is well defined). This gives

$$P(T) = \frac{d}{dT} \left[ \frac{a}{\sqrt{2\pi}} \int_0^T \frac{e^{-a^2/(2v)}}{v^{3/2}} dv \right], \quad (4.7)$$

and so the fundamental theorem of calculus provides the final result, which is

$$P(T) = \frac{a}{\sqrt{2\pi}} \frac{e^{-a^2/(2T)}}{T^{3/2}}. \quad (4.8)$$

If we calculate the expectation value of  $T_a$  (that is, the average time that it will take for the Wiener process to reach the value  $a$ ) then we find that this is

$$\langle T_a \rangle = \frac{a}{\sqrt{2\pi}} \int_0^\infty \frac{e^{-x^2/(2T)}}{\sqrt{T}} dt = \infty. \quad (4.9)$$

This result for  $\langle T_a \rangle$  is reasonable because many paths of the Wiener process go off towards  $-\infty$  and never cross  $a$ .

### 4.3 The stationary auto-correlation function, $g(\tau)$

For a stochastic process, it is often useful to know how correlated the values of the process are at two different times. This will tell us how long it takes the process to forget the value it had at some earlier time. We are therefore interested in calculating the correlation coefficient (see Section 1.4)

$$C_{X(t)X(t+\tau)} = \frac{\langle X(t)X(t+\tau) \rangle - \langle X(t) \rangle \langle X(t+\tau) \rangle}{\sqrt{V(X(t))V(X(t+\tau))}} \quad (4.10)$$

for an arbitrary time difference  $\tau$ . As an illustration we calculate this for the Wiener process. We know already that  $V(W(t)) = t$  and thus  $V(W(t+\tau)) = t + \tau$ . We can calculate the correlation  $\langle W(t)W(t+\tau) \rangle$  in the following way:

$$\begin{aligned} \langle W(t)W(t+\tau) \rangle &= \left\langle \int_0^t dW \int_0^{t+\tau} dW \right\rangle = \left\langle \int_0^t dW \left( \int_0^t dW + \int_t^{t+\tau} dW \right) \right\rangle \\ &= \left\langle \left( \int_0^t dW \right)^2 + \int_0^t dW \int_t^{t+\tau} dW \right\rangle \\ &= \left\langle \left( \int_0^t dW \right)^2 \right\rangle + \left\langle \int_0^t dW \int_t^{t+\tau} dW \right\rangle \end{aligned} \quad (4.11)$$

$$\begin{aligned} &= \langle W(t)^2 \rangle + \left\langle \int_0^t dW \right\rangle \left\langle \int_t^{t+\tau} dW \right\rangle \\ &= \langle W(t)^2 \rangle + \langle W(t) \rangle \langle W(\tau) \rangle = t + 0 = t. \end{aligned} \quad (4.12)$$

And this gives

$$C_{W(t)W(t+\tau)} = \frac{t}{\sqrt{t(t+\tau)}} = \sqrt{\frac{1}{(1 + \tau/t)}}. \quad (4.13)$$

In the above derivation, to get from line (4.11) to (4.12) we used the fact that the random variables  $A = \int_0^t dW$  and  $B = \int_t^{t+\tau} dW$  are *independent*, which implies that their correlation  $\langle AB \rangle$  is just the product of their means,  $\langle A \rangle \langle B \rangle$ . From Eq. (4.13)

we see, as expected, that the Wiener process at time  $t + \tau$  is increasingly independent of its value at an earlier time  $t$  as  $\tau$  increases.

The function  $g(t, t') = \langle X(t)X(t') \rangle$  is often called the *two-time* correlation function or the *auto*-correlation function (“auto” because it is the correlation of the process with *itself* at a later time). If the mean of the process  $X(t)$  is constant with time, and the auto-correlation function,  $g(t, t + \tau) = \langle X(t)X(t + \tau) \rangle$  is also independent of the time,  $t$ , so that it depends only on the *time-difference*,  $\tau$ , then  $X(t)$  is referred to as being “wide-sense” stationary. In this case, the auto-correlation function depends only on  $\tau$ , and we write

$$g(\tau) = \langle X(t)X(t + \tau) \rangle. \quad (4.14)$$

The reason that we call a process whose mean and auto-correlation function are time-independent “wide-sense” stationary, rather than merely “stationary”, is because the latter term is reserved for processes in which the expectation values of products of the process at *any number* of different times only depends on the time-differences. Thus while stationary processes are always wide-sense stationary, wide-sense stationary processes need not be stationary.

The auto-correlation function for a wide-sense stationary process is always symmetric, so that  $g(-\tau) = g(\tau)$ . This is easily shown by noting that

$$g(-\tau) = \langle X(t)X(t - \tau) \rangle = \langle X(t - \tau)X(t) \rangle = \langle X(t)X(t + \tau) \rangle = g(\tau). \quad (4.15)$$

#### 4.4 Conditional probability densities

We now describe the general method for calculating two-time correlation functions. One can always calculate the correlation  $\langle X(t')X(t) \rangle$  at two times  $t'$  and  $t = t' + \tau$ , for some arbitrary process  $X(t)$ , so long as one has the *joint probability density* that the value of the process is  $x$  at time  $t$  and  $x'$  at time  $t'$ . We will write this density as  $P(x, t; x', t')$ . Assume for the sake of definiteness that  $t \geq t'$ . To calculate this joint density we first write it in terms of the conditional probability density  $P(x, t|x', t')$ , which gives

$$P(x, t; x', t') = P(x, t|x', t')P(x', t'). \quad (4.16)$$

The conditional probability is the probability density for  $X$  at time  $t$ , given that  $X$  has the value  $x'$  at time  $t'$ . In fact, we already know how to calculate this, since it is the same thing that we have been calculating all along in solving stochastic differential equations: the solution to an SDE for  $X$  is the probability density for  $X$  at time  $t$ , given that its initial value at  $t = 0$  is  $x_0$ . To obtain the conditional probability in Eq. (4.16), all we need to do is solve the SDE for  $x$ , but this time with the initial time being  $t'$  rather than 0.

As an example, let us do this for the simplest stochastic equation,  $dX = dW$ . Solving the SDE means summing all the increments  $dW$  from time  $t'$  to  $t$ , with the initial condition  $X(t') = x'$ . The solution is

$$X(t) = x' + \int_{t'}^t dW = x' + W(t - t'), \quad (4.17)$$

and this has the probability density

$$P(x, t) = \frac{e^{-(x-x')^2/[2(t-t')]}}{\sqrt{2\pi(t-t')}}. \quad (4.18)$$

This is also the conditional probability for  $X$  at time  $t$ , given that  $X = x'$  at time  $t'$ :

$$P(x, t|x', t') = P(x, t) = \frac{e^{-(x-x')^2/[2(t-t')]}}{\sqrt{2\pi(t-t')}}. \quad (4.19)$$

To calculate the joint probability density we now need to specify the density for  $X$  at time  $t'$ . If  $X$  started with the value 0 at time 0, then at time  $t'$  the density for  $X(t')$  is just the density for the Wiener process, thus

$$P(x', t') = \frac{e^{-x'^2/[2t']}}{\sqrt{2\pi t'}}. \quad (4.20)$$

Using Eqs. (4.19) and (4.20), the joint density is

$$P(w, t; w', t') = P(x, t|x', t')P(x', t') = \frac{e^{-(x-x')^2/[2(t-t')] - x'^2/[2t']}}{2\pi\sqrt{(t-t')t'}}, \quad (4.21)$$

and the correlation function is therefore

$$\langle X(t')X(t) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xx' P(x, t; x', t') dx dx' = t'. \quad (4.22)$$

We then obtain the correlation coefficient between  $X(t')$  and  $X(t)$  by dividing this by the square root of the product of the variances as above.

#### 4.5 The power spectrum

We now return to the Fourier transform introduced in Section 4.1 where we defined the characteristic function. This time we will use a slightly different definition for the Fourier transform: while in Section 4.1 we followed the definition that is standard for characteristic functions, for considering Fourier transforms of signals (functions of time), it is standard to use the following definition. The Fourier transform of the deterministic function  $f(t)$  is

$$F(v) = \int_{-\infty}^{\infty} f(t)e^{-i2\pi vt} dt. \quad (4.23)$$

With this definition the inverse transform is

$$f(t) = \int_{-\infty}^{\infty} F(\nu) e^{i2\pi\nu t} d\nu. \quad (4.24)$$

Note that this time we have also defined the Fourier transform with a minus sign in the exponential. Since either sign is equally good, which one to use is purely a matter of convention. With this definition for the Fourier transform, the Dirac  $\delta$ -function is now the Fourier transform of the constant function  $f(t) = 1$ :

$$\delta(\nu) = \int_{-\infty}^{\infty} e^{-i2\pi\nu t} dt. \quad (4.25)$$

We now recognize that the inverse Fourier transform expresses a function  $f(t)$  as an integral (sum) over the complex exponentials  $e^{i2\pi\nu t}$  with different values of the frequency  $\nu$ . These complex exponentials are complex sinusoids:  $e^{i2\pi\nu t} = \cos(2\pi\nu t) + i \sin(2\pi\nu t)$ . Thus the inverse Fourier transform expresses a function  $f(t)$  as a weighted sum of complex sine waves where the weighting factor is the Fourier transform  $F(\nu)$ . This representation for a function of time is very useful.

All the functions of time that we will be interested in here are real, and when discussing Fourier transforms it is common practice to call them *signals*. Since we are concerned with real signals, you might find it a little strange that in representing a function as a sum over sine waves we choose to use complex exponentials rather than real sine and cosine functions. This choice is a very good one, however, because complex exponentials are easier to manipulate than real sines or cosines by themselves.

If  $f(t)$  is real,  $F(\nu)$  must be such that all the complex parts of the exponentials cancel when the integral over  $\nu$  is taken. If we write  $F(\nu)$  in terms of its magnitude,  $A(\nu)$ , and phase  $\phi(\nu)$ , then we have

$$\begin{aligned} f(t) &= \int_{-\infty}^{\infty} A(\nu) e^{i[2\pi\nu t + \phi(\nu)]} d\nu \\ &= \int_{-\infty}^{\infty} A(\nu) \cos[2\pi\nu t + \phi(\nu)] d\nu + i \int_{-\infty}^{\infty} A(\nu) \sin[2\pi\nu t + \phi(\nu)] d\nu. \end{aligned}$$

Since  $A(\nu)$  and  $\phi(\nu)$  are real, the second term above is purely imaginary, and therefore must be zero. So we have

$$f(t) = \int_{-\infty}^{\infty} A(\nu) \cos[2\pi\nu t + \phi(\nu)] d\nu. \quad (4.26)$$

So we see that  $f(t)$  is the sum of sine waves of different frequencies  $\nu$ , where the amplitude of the wave at frequency  $\nu$  is given by  $A(\nu)$ , and the phase is given by

$\phi(\nu)$ . In terms of  $F(\nu)$ , the amplitude and phase are given by

$$A(\nu) \equiv |F(\nu)| = \sqrt{F^*(\nu)F(\nu)}, \quad (4.27)$$

$$\tan[\phi(\nu)] = \frac{\text{Im}[F(\nu)]}{\text{Re}[F(\nu)]}. \quad (4.28)$$

Here  $F^*(\nu)$  is the complex conjugate of  $F(\nu)$ , and

$$\text{Re}[F(\nu)] = \frac{F(\nu) + F^*(\nu)}{2}, \quad (4.29)$$

$$\text{Im}[F(\nu)] = \frac{F(\nu) - F^*(\nu)}{2i}, \quad (4.30)$$

are, respectively, the real and imaginary parts of  $F(\nu)$ . The Fourier transform is often referred to as the *spectral decomposition* or *complex spectrum* of a signal  $f(t)$ , and  $A(\nu)$  as the (real) spectrum.

The square of the value of a signal at time  $t$  is referred to as the “instantaneous power” of the signal at that time. This comes from the field of electronics, where the signal is usually a voltage difference between two points. In this case the power dissipated by the current flowing between the two points is proportional to the *square* of the voltage difference between the points, and thus to the square of the signal. This relationship between the signal,  $f(t)$ , and the instantaneous power is also true for electromagnetic waves (including light and radio waves) as well as sound waves, if the signal is the amplitude of the wave.

#### 4.5.1 Signals with finite energy

If the instantaneous power of the deterministic signal  $f(t)$  is defined to be  $f^2(t)$ , then the total energy of the signal is the integral of  $f^2(t)$  over all time:

$$E[f(t)] = \int_{-\infty}^{\infty} f(t)^2 dt. \quad (4.31)$$

If the signal has finite duration, then the total energy is finite. Using the definition of the Fourier transform it is not difficult to show that

$$E[f(t)] = \int_{-\infty}^{\infty} f(t)^2 dt = \int_{-\infty}^{\infty} |F(\nu)|^2 d\nu. \quad (4.32)$$

Now, let us say that we pass the signal through a filter that only lets through sinusoids with frequencies between  $\nu_1$  and  $\nu_2$ . Since the Fourier transform gives the amplitude of each sinusoid that makes up a signal, the signal that emerges from the filter (the filtered signal) would have a Fourier transform of

$$G(\nu) = \begin{cases} F(\nu) & \nu \in [\nu_1, \nu_2] \\ 0 & \text{otherwise.} \end{cases} \quad (4.33)$$

The total energy of the filtered signal would then be

$$E[g(t)] = \int_{-\infty}^{\infty} |G(\nu)|^2 d\nu = \int_{\nu_1}^{\nu_2} |F(\nu)|^2 d\nu. \quad (4.34)$$

Thus it makes sense to call  $|F(\nu)|^2$  the spectral energy density of the signal  $f(t)$ .

Once again using the definition of the Fourier transform we find that

$$\begin{aligned} |F(\nu)|^2 &= \int_{-\infty}^{\infty} f(t) e^{-i2\pi\nu t} dt \int_{-\infty}^{\infty} f(t) e^{i2\pi\nu t} dt \\ &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f(t) f(t - \tau) dt \right) e^{-i2\pi\nu\tau} d\tau \\ &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} f(t) f(t + \tau) dt \right) e^{-i2\pi\nu\tau} d\tau \\ &= \int_{-\infty}^{\infty} h(t, \tau) e^{-i2\pi\nu\tau} d\tau \end{aligned} \quad (4.35)$$

To obtain the third line we made the transformation  $t \rightarrow t - \tau$ , and in the last line we defined the function

$$h(t, \tau) = \int_{-\infty}^{\infty} f(t) f(t + \tau) dt. \quad (4.36)$$

Note that  $h(t, \tau)$  is a version of an auto-correlation function, but for deterministic signals. This auto-correlation function is the product of the signal with itself at different times, but this time integrated over time rather than averaged over all realizations. We see that the energy spectrum,  $|F(\nu)|^2$ , is the Fourier transform of this auto-correlation function.

For a stochastic signal,  $x(t)$ , the total average energy in the signal is the average value of the instantaneous power,  $x^2(t)$ , integrated over all time:

$$E[x(t)] = \int_{-\infty}^{\infty} \langle x(t)^2 \rangle dt. \quad (4.37)$$

If the mean of  $x(t)$  is zero for all  $t$ , then of course  $\langle x(t)^2 \rangle$  is equal to the variance of  $x(t)$ .

To calculate the energy spectrum for  $x(t)$ , we must first define what we mean by the Fourier transform of a stochastic process. A stochastic process,  $x(t)$ , has many possible sample paths. Recall from Section 4.1 that we can think of  $x(t)$  as being described by a probability density over the whole collection of possible sample paths. Each one of these sample paths is a function of time, say  $x_\alpha(t)$ , where  $\alpha$  labels the different possible paths. Thus  $x(t)$  is actually a *random function*, whose possible values are the functions  $x_\alpha(t)$ .

Just as  $x(t)$  is a random function, whose values are the sample paths, we define the Fourier transform,  $X(\nu)$ , as a random function whose values are the Fourier transforms of each of the sample paths. Thus the possible values of  $X(\nu)$  are the functions

$$X_\alpha(\nu) = \int_{-\infty}^{\infty} x_\alpha(t) e^{-i2\pi\nu t} dt. \quad (4.38)$$

With this definition of the Fourier transform of a stochastic process, the results above for the energy spectrum are derived in exactly the same way as above for a deterministic process. In particular we have

$$E[x(t)] = \int_{-\infty}^{\infty} \langle |X(\nu)|^2 \rangle dt, \quad (4.39)$$

and

$$\langle |F(\nu)|^2 \rangle = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} g(t, \tau) dt \right] e^{-i2\pi\nu\tau} d\tau, \quad (4.40)$$

where  $g(t, \tau)$  is

$$g(t, \tau) = \langle f(t) f(t + \tau) \rangle. \quad (4.41)$$

#### 4.5.2 Signals with finite power

Many signals we deal with, such as simple sine waves, do not have any specific duration, and we usually think of them as continuing for all time. The energy in such a signal is therefore infinite. This is also true of many stochastic processes – the solution to the Ornstein–Uhlenbeck equation in the long-time limit reaches a stationary solution in which the variance remains constant for the rest of time. For these kinds of signals it is the *average power* of the signal that is finite. For a deterministic process we can define the average power as

$$P = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} f^2(t) dt, \quad (4.42)$$

and this is finite so long as the signal does not blow up at  $\pm\infty$ .

We can define the average power of a stochastic signal in the same way as for a deterministic signal, but this time we take the expectation value of the process, so as to average the power both over time and over all realizations (sample paths) of the process. Thus the average power of a stochastic process is

$$P[x(t)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \langle x^2(t) \rangle dt. \quad (4.43)$$

With the average power defined, it is useful to be able to calculate the *power spectrum*, also called the *power spectral density*,  $S(\nu)$ . This is defined as the power per unit frequency of a sample path of the process, averaged over all the sample paths. To put this another way, we want a function,  $S(\nu)$ , that if we were to put a long sample path of the process  $x(t)$  through a filter that only lets pass frequencies in the range  $\nu_1$  to  $\nu_2$ , the power of the filtered signal  $Y(t)$ , averaged over all sample paths, would be

$$P[Y(t)] = \int_{\nu_1}^{\nu_2} S(\nu) d\nu. \quad (4.44)$$

It turns out that the power spectral density of a wide-sense stationary stochastic process  $x(t)$  is the Fourier transform of the two-time auto-correlation function. Thus

$$S(\nu) = \int_{-\infty}^{\infty} g(\tau) e^{-i2\pi\nu\tau} d\tau, \quad g(\tau) = \langle x(t)x(t+\tau) \rangle. \quad (4.45)$$

This result is called the *Wiener–Khinchin theorem*. Proofs of this theorem can be found in [3] and [15].

It is worth noting that if a process is stationary (see Section 4.3 above) we usually do not need to take the average over all sample paths – the average power in a given frequency range, assuming we have long enough sample paths, is usually the same for all sample paths. This is because the average power is defined as the average over *all* time, and averaging a sample path of a wide-sense stationary process over all time is often the same as averaging over all sample paths. If this is true the process is referred to as being *ergodic*. Not all wide-sense stationary processes are ergodic, however, and this can be very difficult to prove one way or the other. If ergodicity cannot be proven for a given physical process, physicists will often assume that it is true anyway.

#### 4.6 White noise

Consider a function  $f(t)$  whose integral from  $-\infty$  to  $\infty$  is finite. The more sharply peaked  $f$  (that is, the smaller the smallest time interval containing the majority of its energy), then the *less* sharply peaked is its Fourier transform. Similarly, the broader a function, then the narrower is its Fourier transform. Now consider a stochastic process  $x(t)$ . If the auto-correlation function  $g(\tau) = \langle x(t)x(t+\tau) \rangle$  drops to zero very quickly as  $|\tau|$  increases, then the power spectrum of the process  $x(t)$  must be broad, meaning that  $x(t)$  contains high frequencies. This is reasonable, since if a process has high frequencies it can vary on short time-scales, and therefore become uncorrelated with itself in a short time.

We have shown above that the derivative of the Wiener process does not exist. However, there is a sense in which the auto-correlation of this derivative exists, and this fact can be useful as a calculational tool. For the sake of argument let us call the derivative of the Wiener function  $\xi(t)$ . Since the increments of the Wiener process in two consecutive time intervals  $dt$  are independent of each other,  $\xi(t)$  must be uncorrelated with itself whenever the time separation is greater than zero. Thus we must have  $\langle \xi(t)\xi(t+\tau) \rangle = 0$  if  $\tau > 0$ . In addition, if we try to calculate  $\langle \xi(t)\xi(t) \rangle$ , we obtain

$$\begin{aligned} g(0) = \langle \xi(t)\xi(t) \rangle &= \lim_{\Delta t \rightarrow 0} \left\langle \frac{\Delta W}{\Delta t} \frac{\Delta W}{\Delta t} \right\rangle \\ &= \lim_{\Delta t \rightarrow 0} \frac{\langle (\Delta W)^2 \rangle}{(\Delta t)^2} = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \rightarrow \infty. \end{aligned} \quad (4.46)$$

We recall from Chapter 1 that a function which has this property is the delta-function  $\delta(\tau)$ , defined by the rule that

$$\int_{-\infty}^{\infty} f(t)\delta(t) dt = f(0) \quad (4.47)$$

for any smooth function  $f(t)$ . One can think of the delta-function as the limit in which a function with a single peak at  $t = 0$  becomes narrower and taller so that the area under the function is always equal to unity. For example,

$$\delta(t) = \lim_{\sigma \rightarrow 0} \frac{e^{-t^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}. \quad (4.48)$$

Let us see what happens if we assume that  $\xi(t)$  is a noise source with the auto-correlation function  $\langle \xi(t)\xi(t+\tau) \rangle = \delta(\tau)$ . We now try using this assumption to solve the equation

$$dx = gdW. \quad (4.49)$$

Of course we know the solution is just  $x(t) = gW(t)$ . If  $\xi(t)$  exists then we can write the stochastic equation as

$$\frac{dx}{dt} = g\xi(t), \quad (4.50)$$

and the solution is simply

$$x(t) = g \int_0^t \xi(s) ds. \quad (4.51)$$

Now, let us calculate the variance of  $x(t)$ . This is

$$\begin{aligned} V(x(t)) &= \left\langle g^2 \int_0^t \xi(s) ds \int_0^t \xi(v) dv \right\rangle = g^2 \int_0^t \int_0^t \langle \xi(s) \xi(v) \rangle ds dv \\ &= g^2 \int_0^t \int_0^t \delta(s-v) ds dv = g^2 \int_0^t dv = g^2 t, \end{aligned} \quad (4.52)$$

which is the correct answer. Also, we can calculate the two-time auto-correlation function of  $x(t)$ . This is

$$\begin{aligned} \langle x(t)x(t+\tau) \rangle &= \left\langle g^2 \int_0^t \xi(s) ds \int_0^{t+\tau} \xi(v) dv \right\rangle = g^2 \int_0^t \int_0^{t+\tau} \delta(s-v) ds dv \\ &= g^2 \int_0^t dv = g^2 t, \end{aligned} \quad (4.53)$$

which is also correct. So we have been able to obtain the correct solution to the stochastic differential equation by assuming that  $\xi(t) \equiv dW(t)/dt$  exists and has a delta auto-correlation function. This technique will work for any SDE that has *purely additive noise*, but it does not work for SDEs in which the noise  $dW$  multiplies a function of any of the variables. Thus one cannot use it to solve the equation  $dx = xdW$ , for example. Note also that in using the relation  $\langle \xi(t)\xi(t+\tau) \rangle = \delta(\tau)$ , we must use the additional definition discussed in Section 3.8.3: if the peak of the  $\delta$ -function is located at one of the limits of an integral, we must take all its area to be at the lower limit, and none at the upper limit.

Now, since the power spectrum of a process is the Fourier transform of the auto-correlation function, this means that the spectrum of  $\xi$  is

$$S(\nu) = \int_{-\infty}^{\infty} \delta(t) e^{-i2\pi\nu t} dt = 1. \quad (4.54)$$

This spectrum is the same for all values of the frequency  $\nu$ . This means that  $\xi(t)$  contains arbitrarily high frequencies, and thus infinitely rapid fluctuations. It also means that the power in  $\xi(t)$  is infinite. Both of these are further reflections of the fact that  $\xi(t)$  is an idealization, and cannot be truly realized by any real noise source. Because the spectrum contains equal amounts of all frequencies, it is referred to as “white” noise.

The reason that we can assume that  $\xi(t)$  exists, and use this to solve stochastic equations that contain additive noise, is that a differential equation is a filter that does not let through infinitely high frequencies: even though  $\xi(t)$  has infinitely high frequencies, the solution to a differential equation driven by  $\xi(t)$  does not. So long as the bandwidth of the dynamics of the system (that is, the frequencies the differential equation “lets through”) are small compared to the bandwidth of frequencies of a real noise source that drives it, white noise will serve as a good

approximation to the real noise. This is why we can use the Wiener process to model noise in real systems. To derive these results, we would need to discuss linear-system theory, which is beyond our scope here. The question of why we cannot use  $\xi(t)$  (at least as defined above) to model real noise that is not additive is more subtle, and will be discussed in section 5.3.

### Further reading

Further details regarding  $\delta$ -functions, Fourier transforms, filters, and power spectra are given in the excellent and concise text *Linear Systems* by Sze Tan [3]. Fourier transforms and filters in deterministic systems may also be found in, for example, *Signals and Systems* by Alan Oppenheim and Alan Willsky [4]. Further details regarding the power spectral density and its applications are given in *Principles of Random Signal Analysis and Low Noise Design: The Power Spectral Density and its Applications*, by Roy Howard [15]. Techniques for detecting signals in noise are discussed extensively in *Detection of Signals in Noise*, by Robert McDonough and Anthony Whalen [16].

### Exercises

1. Show that
  - (i) If  $f(t)$  is real then the Fourier transform satisfies  $F(\nu) = F(-\nu)^*$ .
  - (ii) If  $F(\nu)$  is the Fourier transform of  $f(t)$ , then the Fourier transform of  $f(at)$  is  $F(\nu/a)/|a|$ .
  - (iii) If  $F(\nu)$  is the Fourier transform of  $f(t)$ , and assuming that  $f(t)$  is zero at  $t = \pm\infty$ , the Fourier transform of  $d^n f/dt^n$  is  $(i2\pi\nu)^n F(\nu)$ .
2. Show that
 
$$\int_{-\infty}^{\infty} f(t)^2 dt = \int_{-\infty}^{\infty} |F(\nu)|^2 d\nu. \quad (4.55)$$

Hint: use the fact that  $\delta(t-\tau) = \int_{-\infty}^{\infty} e^{-i2\pi\nu(t-\tau)} d\nu$ .
3. Calculate the average power of the signal  $f(t) = A \sin(\omega t)$ .
4. Calculate the auto-correlation function of

$$x(t) = \cos(\omega t)x_0 + g \int_0^t \cos[\omega(t-s)]dW(s), \quad (4.56)$$

where  $x_0$  is a random variable, independent of the Wiener process, with mean zero and variance  $V_0$ .

5. Calculate the auto-correlation function of

$$x(t) = e^{-\beta W(t)} W(t). \quad (4.57)$$

6. (i) Calculate the auto-correlation function of the solution to

$$dx = -\gamma x dt + g dW. \quad (4.58)$$

- (ii) Show that in the limit as  $t \rightarrow \infty$  the auto-correlation function is independent of the absolute time  $t$ . In this limit the process therefore reaches a “steady-state” in which it is wide-sense stationary.  
 (iii) In this limit calculate the power spectral density of the process.

7. (i) Calculate the auto-correlation of the solution to

$$dx = -\frac{\beta^2}{2} x dt + \beta x dW. \quad (4.59)$$

- (ii) Show that in the limit as  $t \rightarrow \infty$  the auto-correlation function is independent of the absolute time  $t$ . In this limit the process therefore reaches a “steady-state” in which it is wide-sense stationary.  
 (iii) In this limit calculate the power spectral density of the process.

8. Calculate the auto-correlation function of  $x$  in the long-time limit, where

$$dx = (\omega y - \gamma x) dt, \quad (4.60)$$

$$dy = (-\omega x - \gamma y) dt + g dW. \quad (4.61)$$

## Some applications of Gaussian noise

### 5.1 Physics: Brownian motion

In 1827, a scientist named Robert Brown used a microscope to observe the motion of tiny pollen grains suspended in water. These tiny grains jiggle about with a small but rapid and erratic motion, and this motion has since become known as *Brownian motion* in recognition of Brown’s pioneering work. Einstein was the first to provide a theoretical treatment of this motion (in 1905), and in 1908 the French physicist Paul Langevin provided an alternative approach to the problem. The approach we will take here is very similar to Langevin’s, although a little more sophisticated, since the subject of Ito calculus was not developed until the 1940s.

The analysis is motivated by the realization that the erratic motion of the pollen grains comes from the fact that the liquid is composed of lumps of matter (molecules), and that these are bumping around randomly and colliding with the pollen grain. Because the motion of the molecules is random, the net force on the pollen grain fluctuates in both size and direction depending on how many molecules hit it at any instant, and whether there are more impacts on one side of it or another.

To describe Brownian motion we will assume that there is a rapidly fluctuating force on the molecule, and that the fluctuations of this force are effectively white noise. (The definition of white noise is given in Section 4.6.) Also, we know from fluid dynamics that the pollen grain will experience friction from the liquid. This friction force is proportional to the negative of the momentum of the grain, so that

$$F_{\text{friction}} = -\gamma p = -\gamma m v, \quad (5.1)$$

where  $m$  is the mass of the grain and  $v$  is its velocity. The constant of proportionality,  $\gamma$ , is usually referred to as the damping rate, and is given by  $\gamma = 6\pi\eta a/m$ . Here  $a$  is the diameter of the pollen grain (assumed spherical), and  $\eta$  is the viscosity of the liquid.

We will only treat the motion in one dimension, since we lose nothing important by this restriction, and the extension to three dimensions is simple. The equation of motion for the position of the pollen grain is therefore

$$m \frac{d^2x}{dt^2} = F_{\text{friction}} + F_{\text{fluct}} = -\gamma p + g\xi(t). \quad (5.2)$$

Here  $\xi(t)$  is the white noise process discussed in Chapter 4, with correlation function

$$\langle \xi(t)\xi(t+\tau) \rangle = \delta(\tau), \quad (5.3)$$

and  $g$  is a constant giving the overall magnitude of the fluctuating force. We will determine the value of  $g$  shortly. We can write the equation of motion in vector form, as described in Chapter 2, and this is

$$\frac{d}{dt} \begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} 0 & 1/m \\ 0 & -\gamma \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} + \begin{pmatrix} 0 \\ g\xi(t) \end{pmatrix}. \quad (5.4)$$

Finally we write the equations in differential form, and this gives

$$\begin{pmatrix} dx \\ dp \end{pmatrix} = \begin{pmatrix} 0 & 1/m \\ 0 & -\gamma \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} dt + \begin{pmatrix} 0 \\ g \end{pmatrix} dW, \quad (5.5)$$

where  $dW$  is our familiar Wiener process.

Solving this set of equations is quite simple, because while  $dx$  depends on  $p$ ,  $dp$  does not depend on  $x$ . We first solve the equation for  $p$ , being the Ornstein–Uhlenbeck process

$$dp = -\gamma p dt + g dW. \quad (5.6)$$

Using the techniques in Chapter 3, the solution is

$$p(t) = e^{-\gamma t} p(0) + g \int_0^t e^{-\gamma(t-s)} dW(s). \quad (5.7)$$

The variance of  $p(t)$  is thus

$$V(p(t)) = g^2 \int_0^t e^{-2\gamma(t-s)} ds = \frac{g^2}{2\gamma} (1 - e^{-2\gamma t}). \quad (5.8)$$

Now we know from statistical mechanics that the average kinetic energy,  $\langle E \rangle = \langle p^2/(2m) \rangle$ , of the pollen grain in the steady-state is equal to  $k_B T/2$ , where  $k_B$  is Boltzmann's constant and  $T$  is the temperature of the liquid [17]. Thus in order for our model to agree with what we already know about the physics of a small particle in a fluid, the solution must have a steady-state, and must give the correct value for the steady-state average energy. The first of these means that the probability density for  $p(t)$  should become independent of time as  $t \rightarrow \infty$ . Examining the solution for  $V(p(t))$  above, we see that

$$\lim_{t \rightarrow \infty} V(p(t)) = \frac{1}{2\gamma} \equiv V(p)_{\text{ss}}, \quad (5.9)$$

and thus the variance does reach a steady-state. Now, since the mean value of  $p(t)$  tends to zero as  $t \rightarrow \infty$ , in the steady-state the mean is zero and thus  $\langle p^2 \rangle = V(p(t))$ . For consistency with statistical mechanics we therefore require that

$$\langle E \rangle = \langle p^2/(2m) \rangle = \frac{V(p)}{2m} = \frac{g^2}{4\gamma m} = \frac{g^2}{24\pi\eta d}, \quad (5.10)$$

and this tells us that the strength of the noise must be

$$g = \sqrt{12\pi\eta dkT}. \quad (5.11)$$

We now want to calculate a quantity that we can actually measure in an experiment. To do this we examine the solution for the position of the pollen grain,  $x(t)$ . To obtain this all we have to do is integrate  $p(t)/m$  with respect to time, and this gives

$$\begin{aligned} x(t) &= \int_0^t \frac{p(s)}{m} ds = \frac{p(0)}{m} \int_0^t e^{-\gamma s} ds + \frac{g}{m} \int_0^t \left[ \int_0^s e^{-\gamma(s-s')} dW(s') \right] ds \\ &= \frac{p(0)}{m} \int_0^t e^{-\gamma s} ds + \frac{g}{m} \int_0^t \left[ \int_0^s e^{-\gamma(s-s')} ds' \right] dW(s) \\ &= \frac{1}{m\gamma} (1 - e^{-\gamma t}) p(0) + \frac{g}{m\gamma} \int_0^t (1 - e^{-\gamma s}) dW(s). \end{aligned} \quad (5.12)$$

In the above calculation we switched the order of integration in the double integral. Recall from Section 3.8.3 that this can be done using the usual rules of calculus, a fact which is simple to show by discretizing the integral.

The above expression, Eq. (5.12), is the complete solution for  $x(t)$ . We see from this that the probability for  $x(t)$  is a Gaussian. We can now easily calculate the mean and variance, which are

$$\begin{aligned} \langle x(t) \rangle &= \left\langle \frac{1}{m\gamma} (1 - e^{-\gamma t}) p(0) \right\rangle + \left\langle \frac{g}{m\gamma} \int_0^t (1 - e^{-\gamma s}) dW(s) \right\rangle \\ &= \frac{1}{m\gamma} (1 - e^{-\gamma t}) p(0), \end{aligned} \quad (5.13)$$

$$\begin{aligned} V(x(t)) &= V \left( \frac{1}{m\gamma} (1 - e^{-\gamma t}) p(0) \right) + V \left( \frac{g}{m\gamma} \int_0^t (1 - e^{-\gamma s}) dW(s) \right) \\ &= \frac{g^2}{(m\gamma)^2} \int_0^t (1 - e^{-\gamma s})^2 ds \\ &= \frac{g^2 t}{(m\gamma)^2} + \frac{g^2}{2m^2\gamma^3} [4e^{-\gamma t} - e^{-2\gamma t} - 3]. \end{aligned} \quad (5.14)$$

Langevin realized that the damping (or decay) rate  $\gamma$  was very fast, much faster than the time-resolution with which the particle could be observed in a real

experiment (at least back in 1908). So this means that  $\gamma t \gg 1$ . In this case the variance of  $x(t)$  simplifies to the approximate expression

$$V(x(t)) \approx \frac{g^2 t}{(m\gamma)^2} - \frac{3g^2}{2m^2\gamma^3} = \frac{g^2}{(m\gamma)^2} \left( t - \frac{3}{2\gamma} \right) \approx \frac{g^2 t}{(m\gamma)^2} = \left( \frac{kT}{3\pi\eta d} \right) t. \quad (5.15)$$

Thus, for times at which  $t\gamma \gg 1$ , the variance of the position of the particle will be proportional to time. This is exactly the same behavior as Wiener noise, and because of this Wiener noise is often referred to as Brownian motion. To test if our model for Brownian motion is accurate, one can perform an experiment in which the motion of a pollen grain is observed over time. If we record the amount by which the position changes in a fixed time interval  $T$ , and average the square of this over many repetitions of the experiment, then we can calculate the variance  $V(x(T))$ . This experiment was performed in 1910 by Smoluchowski and others, confirming the scaling of the variance with time. Since that time many more experiments have been performed, and deviations from Wiener noise have been observed at very short time-scales [18].

So how is it that  $x(t)$  can act just like the Wiener process? For this to be the case, it needs to be true that

$$dx \approx \alpha dW, \quad \text{or equivalently} \quad \frac{dx}{dt} \approx \xi(t), \quad (5.16)$$

for some constant  $\alpha$ . Since  $dx/dt = p/m$ , this means that  $p(t)$  is effectively acting like  $\xi(t)$ . The reason for this is that the auto-correlation function of  $p(t)$  decays at the rate  $\gamma$  (see Exercise 4 in Chapter 3), so that if  $\gamma$  is large then the auto-correlation function is sharply peaked, and approximates the auto-correlation function of  $\xi(t)$ , which is a delta-function. Of course, when we say that  $\gamma$  is large, we actually mean that it is large *compared* to something else. In this case the something else is the time resolution (or time-scale),  $\delta t$ , over which the process is observed. By *time-scale* we mean the time that elapses between observations of the process. In this case the process is  $x(t)$ . If we know the value of  $x(t)$  at time  $t = 0$ , and then again at a later time  $t + \delta t$ , then we cannot distinguish it from the Wiener process so long as  $\delta t \gg 1/\gamma$ . So when we say that  $\gamma$  is large enough that we cannot distinguish  $p(t)$  from  $\xi(t)$ , we mean that  $\gamma$  is much larger than  $1/\delta t$ , where  $\delta t$  is the time resolution of our knowledge of  $p(t)$  and  $x(t)$ .

## 5.2 Finance: option pricing

The prices of the stocks (alternatively, shares) of companies exhibit unpredictable fluctuations, and are therefore good candidates for modeling by stochastic

processes. Because of this, stochastic processes have important applications in finance, in particular in the pricing of financial derivatives. A financial derivative is something whose value “derives” in some way from the price of a basic asset, such as a share in a company. An *option* is one such financial derivative. The simplest kind of option, termed a “European” option, is a contract that gives the bearer the right to buy (or to sell) a number of shares at a fixed price (called the “strike price”) at a fixed date in the future. The date in question is called the “maturity time” of the option. Let us assume for the sake of definiteness that the option is an option to buy, called a “call option” (an option to sell is called a “put” option). Now, clearly, if at the maturity time the share price is higher than the strike price, the bearer will exercise his or her option to buy, then immediately sell the shares so obtained at their current value, and pocket the difference! In that case the value of the option at that time is the precisely the amount of money the bearer makes by this transaction. If, on the other hand, at maturity the share price is *lower* than the strike price, the bearer will not exercise the option and the value of the option is zero.

We see that the value of an option at its maturity depends upon the price of the corresponding share (referred to in financial jargon as the “underlying share”) at the time of maturity. The problem of option pricing is to calculate the value of the option well *before* maturity, when the contract is first agreed upon. Clearly the option must be worth something, because once the bearer has the option, they stand to profit from it with some probability, and have nothing to lose by it. But how much should they pay for it? It was Fischer Black and Myron Scholes who first worked out how to calculate the value of an option (in 1973), and it is their analysis that we will describe here. (For this work Fischer and Black were awarded the Bank of Sweden Prize in Memory of Alfred Nobel, often referred to as the “Nobel prize in economics”.)

### 5.2.1 Some preliminary concepts

#### The interest rate

If we put money in a bank account, then the bank pays us interest on a regular basis, and as a result the money accumulates. Each interest payment is a specified fraction of the money in the account. In finance, one always assumes that any money is stored in a bank account, because it makes no sense not to take advantage of interest payments. If the amount of interest that is paid is  $k$  times the present amount, and this interest is paid once each time interval  $\Delta t$ , then the discrete differential equation for the amount of money in the bank account is

$$\Delta M = kM\Delta t. \quad (5.17)$$

As we saw in Chapter 2, the solution to this equation is

$$M(N\Delta t) = \prod_{n=0}^{N-1} (1+k)M(0) = (1+k)^N M(0). \quad (5.18)$$

If we make the accumulation of interest continuous, then we have the differential equation

$$dM = rMdt, \quad (5.19)$$

where  $r$  is called the interest rate. In this case our money now evolves as

$$M(t) = e^{rt} M(0). \quad (5.20)$$

In finance one usually assumes that all money increases exponentially like this at a rate  $r$ , where  $r$  is called the *risk-free* interest rate. It is risk free because the chance of the bank folding and not giving you your money back is assumed to be so small that it can be ignored.

### Arbitrage

A market is inconsistent if a single commodity has two different prices, or two effectively equivalent commodities have different prices (we will give an example of two “equivalent” commodities below). Because of this inconsistency one could buy the commodity at the cheaper price, and sell it at the higher price, making a risk-free profit. Making such a profit out of market inconsistencies is called *arbitrage*. When traders buy at the lower price and sell at the higher price in order to make a profit, then this raises the demand for the lower-priced items, which in turn raises their price. Similarly, selling the higher-priced items increases their supply and lowers their price. The result is that the two prices converge together eliminating the inconsistency. Traders who watch the market and take advantage of arbitrage opportunities, called *arbitrageurs*, thus perform the function of keeping the market consistent. As a result, it is usual in finance to assume that markets are consistent, with any inconsistencies being eliminated quickly as they arise by arbitrageurs.

The assumption that a market is consistent has important consequences for two commodities that are not identical, but which are equivalent, and should therefore have the same price. As an example of this we now consider the concept of a *forward contract*. A forward contract is an agreement to buy a certain item at a specified price,  $F$ , at a specified future date. We will call the present date  $t = 0$ , and the date of the forward contract  $T$ . The question is, given that we know that the current price of the item is  $P$  dollars, what should the agreed price,  $F$ , in the forward contract be? Answering this question is simplest if we assume that you currently own the item. In this case you can sell it for  $P$  dollars now, and write a

forward contract to buy the same item for  $F$  at the future time  $T$ . If you put your  $P$  dollars in the bank, at the future time they will be worth  $P' = e^{rT} P$ . When you buy back the item at time  $T$ , you will pay  $F$  dollars. Thus at time  $T$  you will own the item, and have made a risk-free profit of

$$\text{money for nothing} = P' - F = e^{rT} P - F. \quad (5.21)$$

This is positive if  $F < e^{rT} P$ . We can alternatively cast this ability to lock in a risk-free profit as an arbitrage situation taking advantage of a market inconsistency. Here there are two things that are equivalent both at the present time, and at the future time  $T$ . The first is  $P$  dollars of money at the present time, and a forward contract to buy an item for a “forward price”  $F$  dollars at a future time  $T$ . The second is the item whose present price is  $P$  dollars. Both of these are equivalent to having the item now, and having the item at the future time. Thus, if these two things do not have the same price at the future time (being the item plus zero dollars) then there is an arbitrage opportunity.

If alternatively the forward price  $F$  is greater than  $e^{rT} P$ , then one can make money from the arbitrage opportunity by being the person who is on the selling end of the forward contract. In this case we borrow  $P$  dollars, use it to buy the item, and then write a forward contract to sell the item for  $F$  dollars at time  $T$ . Assuming we can borrow money at the risk-free interest rate, then at time  $T$  we must pay back  $e^{rT} P$  dollars, and we receive  $F$  dollars from the sale of the item. This time the profit we make is  $F - e^{rT} P$ .

So if we assume that there are no arbitrage opportunities, this tells us that the forward price one should set for a commodity on a forward contract is the present price multiplied by  $e^{rT}$ . That is,  $F = e^{rT} P$ . If the forward contract is made out for a different forward price,  $G$ , then the buyer of the contract (the person who has agreed to buy the commodity) can make  $M = Pe^{rT} - G$  dollars at time  $T$ . We can make this second situation consistent with no-arbitrage by making the buyer pay  $e^{-rT} M = P - e^{-rT} G$  dollars for the contract. So this means that a forward contract that is *not* written with a forward price  $F = e^{rT} P$  is itself worth something. It has a value itself – let us call this  $\mathcal{F}$  – and so must be bought for this price by the buyer of the contract. Note also that once the forward contract has been written with a certain forward price,  $G$ , as time goes by the current price of the commodity will usually change. This means that a forward contract that was initially worth nothing itself, so that  $\mathcal{F} = 0$ , will be worth something (which could be negative) as time goes by. Because of this, forward contracts are bought and sold up until the date  $T$ . Forward contracts that are bought and sold like this are called “futures”.

To find the price of an option (for example, a contract that gives the bearer an option to buy or sell a fixed amount of shares at given price at a future date) one similarly uses a “no-arbitrage” argument, and this is the subject of the next section.

### 5.2.2 Deriving the Black–Scholes equation

To begin with we need to define the term “portfolio”, with which most people are probably already familiar: a portfolio is a collection of assets owned by someone, and which for whatever reason we group together as single unit. A portfolio may consist of a variety of shares in various companies, or it could be a collection of more general assets including shares, money in bank accounts, real estate or even contracts such as options.

Now, let us denote the price of a single share of a company by  $S(t)$ . We have written this as a function of time, since the price of the share will vary as time goes by. We know from experience that the prices of shares fluctuate in an erratic and apparently random fashion. Thus it seems reasonable that we should describe the price of a share as a stochastic process. The question is, which stochastic process? It seems reasonable that the amount by which a share is likely to fluctuate randomly in a given time interval is proportional to its current value. That is, that a share worth 1 dollar would be as likely to move by 10 cents over a period of a week, as a share worth 10 dollars would be to move by 1 dollar. While this may not be exactly true, it is very reasonable because, after all, money is an arbitrary measure of value, and thus the *absolute* value of a share is not a particularly meaningful quantity. We will therefore assume that the random part of the change in  $S$ ,  $dS$ , in time interval  $dt$  is given by

$$dS_{\text{random}} = \sigma S dW, \quad (5.22)$$

where  $dW$  is our familiar Wiener process, and  $\sigma$  is a constant. We also know from looking at history that share prices, on average, grow exponentially with time, just like money in a bank account. We will therefore choose the deterministic part of the increment in the share to be one that generates exponential growth. Thus

$$dS = \mu S dt + \sigma S dW = (\mu dt + \sigma dW)S. \quad (5.23)$$

Here  $\mu$  is referred to as the *expected rate of return* of the share, and  $\sigma$  is called the *volatility*. The volatility determines how large the random fluctuations are. This is the model that Black and Scholes used to describe the time evolution of a share. Note that it is the time-independent multiplicative stochastic equation for which we found the solution in Chapter 3. The advantage of this model is that it is simple, but we now know, from detailed statistical analyses of historical share prices, that it is not quite right. This is mainly because the assumption that the random fluctuations are Gaussian is not quite right, and there is considerable research devoted to devising more accurate models of the stock market. However, it is useful because

it is sufficiently simple that we can derive fairly simple expressions for the prices of some kinds of options, and it is sufficiently realistic that it is still used by traders as a practical tool to calculate option prices.

To work out the price of a call option,  $C$ , we are going to construct two equivalent portfolios, one that includes the option, and another whose time dependence we know. From the assumption of no arbitrage, we will then be able to determine the value of the option. To make the first portfolio, we will need to be able to own a *negative* amount of an asset. So what does this mean? This means that when the value of the asset goes up by an amount  $x$ , the value of our portfolio goes *down* by the same amount. To make a portfolio like this we do the following. We borrow the asset off somebody, with an agreement to give it back at a time  $T$  later, and then immediately sell it (this is called short-selling). If its current price is  $P$ , then we now have  $P$  dollars. Our portfolio consists of a debt of the asset to the person we borrowed it off, and  $P$  dollars in the bank, being the current price of the asset. Now let us see what the value of our portfolio is at time  $T$  if the price of the asset has gone up. To be able to give the asset back to the person we borrowed it off, we must buy the asset back. If the new price of the asset is  $P + x$ , then we have to spend  $P + x$  dollars. Thus the money we have after we buy back the asset is *reduced* by  $x$  dollars. What is the total value of our portfolio at time  $T$ ? The money we have has gone up by the interest we earned, which is  $(e^{rT} - 1)P$ , and has gone *down* by  $x$  dollars. Thus, the portfolio behaves just as if we have positive  $P$  dollars in the bank, and a *negative* amount of the asset.

To work out the price of an option we make a portfolio out of the option and a carefully chosen amount of the asset, and show that this portfolio has no risk. Since it has no risk, it must be equivalent to having money in the bank. To construct this special portfolio, consider an option that gives the bearer the right to buy or sell an asset,  $S$ , whose price is  $S$ . We say that the option is “written” on the asset, and that  $S$  is the “underlying asset”, or the “underlier” for the option. We now make the assumption that the value of the option,  $C$ , (whatever it is) will depend only on the value of the underlying asset,  $S$ , and time. Thus we can write  $C = C(S, t)$ . We will also assume that the value of the option is a smooth function of the asset price and of time, so that the derivatives  $\partial C / \partial S$ ,  $\partial^2 C / \partial S^2$ , and  $\partial C / \partial t$  exist. (We will find that this is true.) Now we construct a portfolio consisting of the option and an amount of the underlying asset  $S$ . The value of the portfolio is

$$\Pi = C + \alpha S, \quad (5.24)$$

where  $\alpha$  is the amount of the asset. We want to work out the change in the value of the portfolio when we increment time by  $dt$ . Since we know the stochastic differential equation for  $S$ , we can work out the stochastic differential equation for

$\Pi$ . We have

$$\begin{aligned}
 d\Pi &= dC + \alpha dS \\
 &= C(S + dS, t + dt) - C(S, t) + \alpha dS \\
 &= C(S, t) + \frac{\partial C}{\partial S} dS + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} (dS)^2 + \frac{\partial C}{\partial t} dt - C(S, t) + \alpha dS \\
 &= \left[ \frac{\partial C}{\partial S} + \alpha \right] dS + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} (dS^2) + \frac{\partial C}{\partial t} dt \\
 &= \left[ \frac{\partial C}{\partial S} + \alpha \right] (\mu S dt + \sigma S dW) + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} (S^2 \sigma^2 dt) + \frac{\partial C}{\partial t} dt \\
 &= \left[ \mu S \left( \frac{\partial C}{\partial S} + \alpha \right) + \frac{\partial C}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \right] dt + \sigma S \left[ \frac{\partial C}{\partial S} + \alpha \right] dW.
 \end{aligned} \tag{5.25}$$

Of course, we cannot solve this stochastic equation, since we don't (yet) know what  $\partial C/\partial S$ ,  $\partial^2 C/\partial S^2$  and  $\partial C/\partial t$  are. However, we observe from this equation that if we chose the amount of the asset to be

$$\alpha = -\frac{\partial C}{\partial S}, \tag{5.26}$$

then the equation for  $dS$  becomes completely deterministic! That is, if we choose the right amount of the asset, any stochastic change in the price of the option is exactly offset by the change in the price of the asset. In this case the value of the portfolio  $\Pi$  has no random component, and therefore is just like having  $\Pi$  dollars in the bank. Since this is equivalent to money, in order to avoid arbitrage opportunities, the change in the value of this portfolio in the time interval  $dt$  must be the same as that of money. Now recall that the "equation of motion" for money  $M$  is

$$dM = rMdt, \tag{5.27}$$

where  $M$  is the "risk-free" interest rate. We can therefore conclude that when we set  $\alpha = -\partial C/\partial S$ , then  $d\Pi$  should also obey the equation

$$d\Pi = r\Pi dt. \tag{5.28}$$

Equating the right-hand sides of Eq. (5.28) and Eq. (5.25), we have

$$r\Pi dt = \left[ \frac{\partial C}{\partial t} + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} S^2 \sigma^2 \right] dt. \tag{5.29}$$

Dividing both sides by  $dt$ , replacing  $\Pi$  with  $C + \alpha S = C - (\partial C/\partial S)S$ , and rearranging, gives

$$\frac{\partial C}{\partial t} = rC - rS \frac{\partial C}{\partial S} - \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2}. \tag{5.30}$$

This is a partial differential equation for the option price  $C$ , and is called the *Black–Scholes* equation. To obtain the option price  $C$ , one solves this equation.

### 5.2.3 Creating a portfolio that is equivalent to an option

To derive the Black–Scholes equation, we used the option and the underlying asset to create a portfolio that had no stochastic part, and was therefore equivalent to money. However, we can alternatively construct a portfolio from an amount of money and some of the underlying asset that is equivalent to the option. This allows us to artificially create the option. To do this we need to construct a portfolio whose change in time interval  $dt$  is the same as that of the option. To work out what this portfolio needs to be, all we have to do is rearrange the portfolio with value  $\Phi$  that we have already constructed. We know that

$$dC = d\Pi - \alpha dS \tag{5.31}$$

and we know that when  $\alpha = -\partial C/\partial S$ , then  $d\Pi$  is just like having money. So let's make a portfolio out of  $M$  dollars and  $-\alpha$  of the asset  $S$ . Denoting the value of this portfolio by  $\Phi$ , we have

$$\Phi = M - \alpha S \tag{5.32}$$

and

$$d\Phi = dM - \alpha dS. \tag{5.33}$$

Now, the change in this portfolio,  $d\Phi$ , will be equal to  $dC$  when we choose the amount of money to be  $M = \Pi = C - \alpha S$ . So to construct a portfolio that will change its value in the next time increment  $dt$  by exactly the same amount as the option, we first calculate the current value of the option  $C$  by solving the Black–Scholes equation. We then purchase  $\alpha S$  worth of the asset and put  $M = C - \alpha S$  dollars in the bank (so as to earn the risk-free interest rate). Note that to do this we have to use a total of  $C$  dollars, which is just the current value of the option. This is expected, because in order to make a portfolio that is equivalent to the option  $C$ , it should have the value of  $C$  at every time, and we should therefore have to spend this amount of money to obtain it. If we now wait a time  $dt$ , we know that the new value of the portfolio,  $\Phi(t + dt)$  still has the same value as the option,  $C(t + dt)$ , because the change in the two are the same.

To make our portfolio continue to match the value of the option after the next time-step  $dt$ , so that  $\Phi(t + 2dt) = C(t + 2dt)$ , we must change the portfolio somewhat. The reason is that to make this work we had to chose  $\alpha = -\partial C/\partial S$ , and when we move one time-step to  $t + dt$ , not only will the value of  $C$  change, but the value of the derivative  $\partial C/\partial S$  will change as well. Now that we have moved one time-step, we will therefore have to change the value of  $\alpha$ , which means that we will have to change the amount of stock we have in our portfolio. Since we know that our portfolio has the value of the option at the present time,  $C(t + dt)$ , and since we know that to create the correct portfolio we need precisely  $C(t + dt)$  dollars, we do not need to add or subtract any money from the portfolio, all we have to do is to “rebalance” the portfolio so that the portion of the total value that is invested in the underlying asset is  $(\partial C(t + dt)/\partial S)S$ . Of course, for this to work, there needs to be no extra costs associated with the act of purchasing and selling shares (so-called “transaction costs”). If there are some transaction costs, then it will cost us more to create a portfolio equivalent to the option than the “true” value of the option, because we would have to spend this extra money as time went by to periodically rebalance the portfolio.

It is this process of artificially creating an option that allows financial institutions to sell options to people. They sell the option to someone for its current value  $C$ , and then use the money to construct the portfolio that is equivalent to the option. They then rebalance the portfolio at intervals to keep it equivalent to the option (they can't do this exactly, but they do it well enough that their risk remains within acceptable bounds). Then, at the time the option matures, if the option is worth something (if it is “in the money”) the institution will have exactly this amount in the portfolio, and can give it to the owner of the option. Since constructing the artificial option costs the institution no more than the price that the owner paid for the option, the institution is taking on only minimal risk when it sells the option. However, because of transaction costs, and the work that the institution does in rebalancing the portfolio, it will charge a bit more for the option than its “true” value  $C$ .

#### 5.2.4 The price of a “European” option

We now discuss the solution of the Black–Scholes equation for European options. The Black–Scholes equation is a *partial* differential equation for the option price  $C$ . This means that  $C$  is a function of more than one variable (in this case the two variables  $S$  and  $t$ ) and that the differential equation contains partial derivatives with respect to all of these variables. In this case we call  $C$  the *dependent* variable, and  $S$  and  $t$  the *independent* variables. The differential equations that we have dealt with so far in this course have involved derivatives (actually, differentials) of only

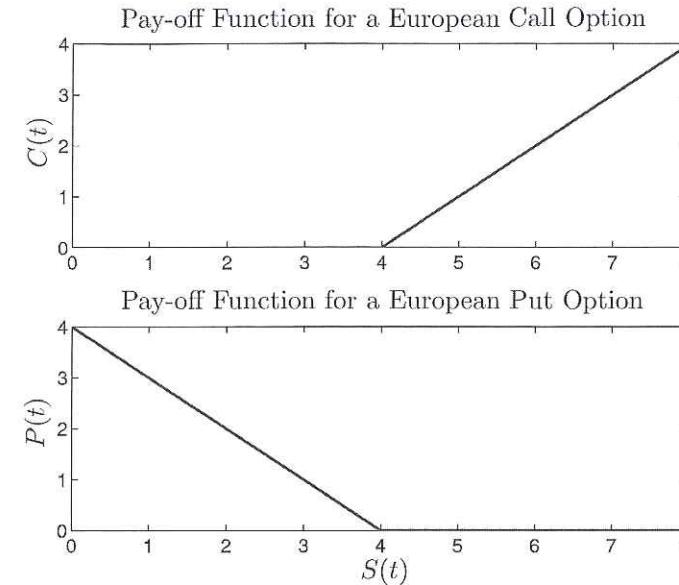


Figure 5.1. The pay-off functions for European call and put options with the strike price  $E = 4$ .

one independent variable, time. In this case the solution for  $y$  at time  $t$  depends on the initial value of  $y$ , or, more generally, on the value of  $y$  specified at any single instant of time.

To obtain a solution to the partial differential equation for  $C$ , where the independent variables are  $t$  and  $S$ , we must specify the value of  $C$  at an initial time  $t = 0$  (or, in fact, at *any* single time  $t = T$ ), and for *all* values of  $S$  at this initial time. In this case, we call these specified values the *boundary* conditions, rather than just the *initial* conditions. Thus to obtain the solution for the value,  $C$ , of an option at the present time, we need to know the value of the option at *some* specified time. Fortunately it is easy for us to work out the value of the option at the maturity (or expiry) time  $T$ .

The value of the option at the time of maturity, as a function of the share price, is called the *pay-off* function. For a European call option for which the strike price is  $E$ , then at the time the option matures, it is worth  $S(T) - E$  if  $S(T) > E$  (where  $S$  is the share price, and  $T$  is the maturity time), or zero if  $S(T) < E$ . We can therefore write the pay-off function for a European call as  $f(S) = \max(S(T) - E, 0)$ . For a European put (which entitles the bearer to sell the shares for a fixed price  $E$ ), the pay-off function is  $F(S) = \max(E - S(T), 0)$ . We plot these pay-off functions in Figure 5.1.

In addition to specifying the values of  $C$  at a particular time, we also need to specify the values it takes at the boundaries (the extreme values) of the  $S$  variable.

These are  $S = 0$  (the shares have no value) and  $S \rightarrow \infty$  (being the limit in which the share price is much larger than the strike price). By examining the stochastic equation for  $S$ , we see that when  $S = 0$  it will remain zero for all later times. This means that when  $S = 0$  the value of the call option is zero. When  $S$  tends to infinity then the value of the call option also tends to infinity. For the put option we have to think a little bit harder. When  $S = 0$  at some time  $t$ , since we know it will remain zero, we know that the value of the put option at maturity will be the strike price  $E$ . If we know that we will have  $E$  dollars at time  $T$ , then this is the same as having  $e^{-(T-t)}E$  dollars in the bank at time  $t$ . Thus, on the  $S = 0$  boundary, the put option has the value  $P(0, t) = e^{-(T-t)}E$ . When  $S \rightarrow \infty$ , then  $S > E$  so that the value of the put is 0.

Armed with the values of the European call and put options at the final time  $T$ , and on the boundaries  $S = 0$  and  $S \rightarrow \infty$ , one can solve the Black–Scholes equation to obtain the value of these options at any earlier time  $t$ , including the current time  $t = 0$  (being the time at which the option is sold). We will not discuss the method by which the Black–Scholes equation is solved, but merely present the solutions. The solution for the European call is

$$C(S, t) = SD(x) - Ee^{-r(T-t)}D(y), \quad (5.34)$$

where  $D$  is the distribution function for a Gaussian random variable:

$$D(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-z^2/2} dz, \quad (5.35)$$

and  $x$  and  $y$  are given by

$$x = \frac{\log(S/E) + (r + \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} \quad (5.36)$$

and

$$y = \frac{\log(S/E) + (r - \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}. \quad (5.37)$$

The value of the European put is

$$P(S, t) = -SD(-x) + Ee^{-r(T-t)}D(-y). \quad (5.38)$$

It is important to note that the Black–Scholes equation is not specific to options. To derive the Black–Scholes equation we used only the fact that the value of  $C$  was a function of the price of the share  $S$  and time  $t$ . Thus *any* portfolio whose value  $V$  is a function only of  $S$  and  $t$  will satisfy the Black–Scholes equation. The value of the portfolio at all times (being the solution of the Black–Scholes equation) is therefore determined completely by the value of the portfolio at a specified final time, and at the other “boundaries”  $S = 0$  and  $S \rightarrow \infty$ . Thus *any* portfolio that

consists only of money, some amount  $S$  of shares, and some amount of options on these shares will satisfy the Black–Scholes equation. As a result, two portfolios of this kind whose values are equal to each other for all  $S$  at a *single* time  $T$ , will have the same value for *all* time, and therefore be *equivalent* portfolios.

### 5.3 Modeling multiplicative noise in real systems: Stratonovich integrals

In the example of Brownian motion discussed above, the noise did not depend on the state of the particle (the position or velocity of the particle). Noise that is independent of the state of a system is said to be “additive”. How do we describe a fluctuating force that is proportional to the position of the particle? Having already treated multiplicative noise such as this in Chapter 3, we are tempted to say that the noise would be described by adding a term proportional to  $xdW$  to the equation for  $dp$ . However, this is not quite correct – there is a subtlety here that must be addressed to get the description right.

The reason for the subtlety is that Wiener noise, the basis of our mathematical treatment of noise, is an idealization. It assumes that in each infinitesimal time interval (no matter how small) the increment due to the noise is *different* (and independent of) the increment in the previous interval. This means that the noise has fluctuations that are infinitely rapid! No physical process can really have this property. To see this another way, consider the average acceleration of the Brownian particle in a given small time interval  $\Delta t$ . This is

$$\bar{a}(\Delta t) = \frac{\Delta p}{\Delta t} = -\gamma p + \beta \frac{\Delta W}{\Delta t}. \quad (5.39)$$

Now, since the standard deviation of  $\Delta W$  is  $\sqrt{\Delta t}$ , the last term is of order  $1/\sqrt{\Delta t}$ , and thus goes to infinity as  $\Delta t$  goes to zero. Thus the force on the particle is infinite (but fluctuating infinitely fast so that the average over a finite time is finite).

While all real forces will be finite, and thus cannot truly be Wiener noise, as discussed in Section 4.6 the Wiener process is a good approximation to a real fluctuating force so long as this force has fluctuations in a frequency range that is broad compared to the frequency range of the motion of the system. This is why the stochastic differential equations we have been studying are useful for describing physical systems. However, it turns out that there is more than one way to define a stochastic difference equation driven by random increments  $\Delta W$ , and thus more than one way to take the continuum limit to obtain a stochastic differential equation.

The general Ito difference equation is

$$\Delta x = f(x, t)\Delta t + g(x, t)\Delta W. \quad (5.40)$$

Now, if we define  $x_n \equiv x(n\Delta t)$ ,  $t_n \equiv n\Delta t$  and the first random increment  $\Delta W$  as  $\Delta W_0$ , then we can write the equation as

$$\begin{aligned} x_n &= x_{n-1} + \Delta x_{n-1} \\ &= x_{n-1} + f(x_{n-1}, t_{n-1})\Delta t + g(x_{n-1}, t_{n-1})\Delta W_{n-1}. \end{aligned} \quad (5.41)$$

The general solution (obtained by repeatedly adding  $\Delta x_n$  to the initial value of  $x$ ,  $x_0 \equiv x(0)$ ) is given by

$$x_N = x(N\Delta t) = x_0 + \sum_{n=0}^{N-1} f(x_n, t_n)\Delta t + \sum_{n=0}^{N-1} g(x_n, t_n)\Delta W_n. \quad (5.42)$$

This expression is not very helpful for finding an explicit solution for  $x(N\Delta t)$ , because to do we so we would have to repeatedly substitute in the solutions for  $x_{N-1}$ ,  $x_{N-2}$ , etc. This would give an unwieldy expansion, ultimately with an infinite number of terms if we tried to take the continuum limit  $N \rightarrow \infty$ . However, this does show us how the Ito integral, defined as

$$\int_0^t g(x(t), t) dW \equiv \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} g(x_n, t_n)\Delta W_n, \quad (5.43)$$

appears in the solution for  $x(t)$ . The subtle point alluded to above is that there is more than one way to define a stochastic integral of the Wiener process. Another way, referred to as the *Stratonovich* integral, is

$$\begin{aligned} \oint_0^t g(x(t), t) dW &\equiv \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} g\left(\frac{x[(n+1)\Delta t] + x[n\Delta t]}{2}, n\Delta t\right) \Delta W_n \\ &\equiv \lim_{N \rightarrow \infty} \sum_{n=1}^N g\left(\frac{x_{n+1} + x_n}{2}, t_n\right) \Delta W_n, \end{aligned} \quad (5.44)$$

where we have written the integral with an ‘‘S’’ so as not to confuse it with the Ito integral. Now, if  $dW$  was the increment of a nice smooth function of time the Ito and Stratonovich integrals would be equal, but this is not the case. Hence we appear to have an ambiguity when defining the integral of the Wiener process – which one should we use?

Before we answer that question, we will calculate the precise relationship between the two kinds of stochastic integrals, the Ito integral – which is what we have been using in solving our stochastic equations – and the Stratonovich integral. To understand where the difference comes from, note that in the Ito integral, in the  $n$ th time-step, it is the value of  $g(x, t)$  at the *start* of the time-step that multiplies the increment  $\Delta W_n$ . The value of  $g$  at this time is a function of all the *previous*

increments,  $\Delta W_{n-1}$ ,  $\Delta W_{n-2}$ , etc., but *not* of the new increment  $\Delta W_n$ . Thus  $(\Delta W_n)^2$  does not appear in the  $n$ th term. However, in the Stratonovich integral, the  $n$ th term contains the value of  $g$  at the *end* of the interval for that time-step. This means that in each time-step, the Stratonovich integral contains an extra term proportional to  $(\Delta W_n)^2$ . As we mentioned above, this would not matter if  $W(t)$  was a sufficiently smooth function (that is, differentiable), because  $(\Delta W_n)^2$  would tend to zero in the continuum limit. However, the square of the Wiener increment does not vanish in this limit, but is instead equal to  $dt$ . It is this that results in the difference between the Ito and Stratonovich integrals. We now calculate precisely what this difference is.

We first define  $x_n$  as the solution to the Ito equation

$$x_n = x_{n-1} + f(x_{n-1}, t_{n-1})\Delta t + g(x_{n-1}, t_{n-1})\Delta W_{n-1}. \quad (5.45)$$

By referring to Eq. (5.45) as an Ito equation we mean that  $x_n$  is the solution to this stochastic equation where the solution is given by using the Ito definition of a stochastic integral, defined in Eq. (5.43). Next we note that the  $n$ th term in the Stratonovich integral is

$$g\left(\frac{x_{n+1} + x_n}{2}, t_n\right) \Delta W_n = g\left(x_n + \frac{\Delta x_n}{2}, t_n\right) \Delta W_n. \quad (5.46)$$

We now expand  $g$  as a power series about the point  $(x_n, t_n)$ . To make the notation more compact we define  $g_n \equiv g(x_n, t_n)$ . Taking this expansion to second order gives

$$g\left(x_n + \frac{\Delta x_n}{2}, t_n\right) = g_n + \left(\frac{\Delta x_n}{2}\right) \frac{\partial g_n}{\partial x} + \left(\frac{\Delta x_n}{2}\right)^2 \frac{1}{2} \frac{\partial^2 g_n}{\partial x^2}. \quad (5.47)$$

As always, the reason we take this to second order is because  $\Delta x_n$  has a  $\Delta W_n$  lurking in it. Now we use the fact that  $\Delta x_n = f(x_n, t_n)\Delta t + g(x_n, t_n)\Delta W_n$ , and remembering that  $\Delta W^2 = \Delta t$  this gives

$$g\left(x_n + \frac{\Delta x_n}{2}, t_n\right) = g_n + \left(\frac{f_n}{2} \frac{\partial g_n}{\partial x} + \frac{g_n^2}{4} \frac{\partial^2 g_n}{\partial x^2}\right) \Delta t + \left(\frac{g_n}{2} \frac{\partial g_n}{\partial x}\right) \Delta W_n. \quad (5.48)$$

Now we have expressed the  $g$  that appears in the  $n$ th term in the Stratonovich integral entirely in terms of the  $g_n$  that appears in the Ito integral. Substituting this into the discrete sum for the Stratonovich integral will immediately allow us to write this sum in terms of that for the Ito integral. The result is

$$\sum_{n=0}^{N-1} g\left(\frac{x_{n+1} + x_n}{2}, t_n\right) \Delta W_n = \sum_{n=0}^{N-1} g_n \Delta W_n + \sum_{n=0}^{N-1} \frac{g_n}{2} \frac{\partial g_n}{\partial x} \Delta t. \quad (5.49)$$

Note that in doing this we have dropped the terms proportional to  $\Delta t \Delta W_n$ . Recall that this is fine because, like  $(\Delta t)^2$ , they vanish in the continuum limit. The first sum on the right hand side is the familiar sum for the Ito integral, so taking the continuum limit we have

$$\oint_0^t g(x(t), t) dW = \int_0^t g(x(t), t) dW + \frac{1}{2} \int_0^t \frac{\partial g(x(t), t)}{\partial x} g(x(t), t) dt. \quad (5.50)$$

We now see that if we define our solution to the stochastic equation

$$dx = f(x, t)dt + g(x, t)dW \quad (5.51)$$

using the Stratonovich integral instead of the Ito integral, then this solution, which we will call  $y(t)$ , would be

$$\begin{aligned} y(t) &= y(0) + \int_0^t f(y(t), t) dt + \oint_0^t g(y(t), t) dW \\ &= y(0) + \int_0^t \left[ f(y(t), t) + \frac{g(y(t), t)}{2} \frac{\partial g(y(t), t)}{\partial y} \right] dt + \int_0^t g(y(t), t) dW. \end{aligned}$$

But this would be the solution we get if we used the Ito integral to solve the stochastic equation

$$dy = \left[ f + \frac{g}{2} \frac{\partial g}{\partial y} \right] dt + gdW. \quad (5.52)$$

So we see that using the Stratonovich integral as the solution to a stochastic differential equation is the same as solving the equation using our usual method (the Ito integral) but changing the deterministic term  $f$  by adding  $(g/2)\partial g/\partial x$ .

Naturally we can also derive the relationship between Ito and Stratonovich stochastic vector equations. Let us write the general vector Stratonovich equation

$$d\mathbf{x} = \mathbf{A}(\mathbf{x}, t)dt + \mathbf{B}(\mathbf{x}, t)d\mathbf{W} \quad (5.53)$$

explicitly in terms of the elements of  $\mathbf{A}$  and  $\mathbf{B}$ . Denoting these elements as  $A_i$  and  $B_{ij}$  respectively, the stochastic equation is

$$dx_i = A_i(\mathbf{x}, t)dt + \sum_{j=1}^N B_{ij}(\mathbf{x}, t)dW_j \quad (\text{Stratonovich}), \quad (5.54)$$

where the  $x_i$  are the elements of  $\mathbf{x}$ , and the  $dW_j$  are the mutually independent Wiener noises that are the elements of  $d\mathbf{W}$ . If Eq. (5.54) is a Stratonovich equation, then the equivalent Ito equation is

$$dx_i = \left( A_i + \frac{1}{2} \sum_{j=1}^N \sum_{k=1}^N B_{kj} \frac{\partial B_{ij}}{\partial v_k} \right) dt + \sum_{j=1}^N B_{ij} dW_j \quad (\text{Ito}). \quad (5.55)$$

So why have we been studying the Stratonovich integral? The reason is that when we consider multiplicative noise in a real physical system, it is the Stratonovich integral rather than the Ito integral that appears naturally: it turns out that if we have a real fluctuating force, and take the limit in which the bandwidth of the fluctuations becomes very broad (compared to the time-scale of the system), it is the Stratonovich integral which is the limit of this process, *not* the Ito integral. To treat multiplicative noise in physical systems, we should therefore use the following procedure.

#### Procedure for describing multiplicative noise in physical systems

When we write down a stochastic equation describing the dynamics of a system driven by physical noise whose magnitude depends on the system's state, this is a Stratonovich stochastic equation. Because Ito equations are much simpler to solve than Stratonovich equations, we transform the Stratonovich equation to an Ito equation before proceeding. Then we can use all the techniques we have learned in Chapters 3 and 4. If the noise is purely additive, then the Stratonovich and Ito equations are the same, and so no transformation is required.

#### Further reading

Recent work on Brownian motion can be found in [18] and references therein. An application of Brownian motion to chemical transport in cells is discussed in [19]. There are numerous books on option pricing and financial derivatives. Some use sophisticated mathematical language, and some do not. One that does not is *The Mathematics of Financial Derivatives: A Student Introduction* by Wilmott, Howison, and Dewynne [20]. Further details regarding Stratonovich equations can be found in the *Handbook of Stochastic Methods*, by Crispin Gardiner [23], and *Numerical Solution of Stochastic Differential Equations*, by Kloeden and Platen [21].

#### Exercises

1. The current time is  $t = 0$ , and the risk-free interest rate is  $r$ . Assume that there is a commodity whose current price is  $P$ , and that a forward contract for the date  $T$  is written for this commodity with forward price  $F = e^{rT} P$ . Determine an expression for the value of the forward contract,  $\mathcal{F}(t)$ , as a function of time, given that the price of the commodity as a function of time is  $P(t)$ .
2. (i) Calculate  $\partial C / \partial S$  for the European call option. (ii) Show that the expression for the European call option (Eq. (5.34)) is a solution of the Black–Scholes equation.

3. (i) Show that a very simple solution of the Black–Scholes equation is

$$V(S, t) = aS + be^{rt}, \quad (5.56)$$

where  $a$  and  $b$  are constants. (ii) What portfolio does this solution represent?

4. In Section 5.2.2 we explained how one could have a *negative* amount of an asset in one's portfolio. (Having a negative amount of an asset is called "shorting" the asset.) Options are also assets, so we can have a portfolio that is

$$V(S, t) = C - P, \quad (5.57)$$

where  $P$  is a put option and  $C$  is a call option. (i) If the call and put have the same maturity time  $T$  and strike price  $E$ , draw the pay-off function for the portfolio  $V$ .

(ii) Plot the value of the portfolio

$$U(S, t) = S - Ee^{r(t-T)}, \quad (5.58)$$

at time  $T$ . (iii) How is the value of  $U$  at time  $T$  related to the value of  $V$  at time  $T$ , and what does this tell you about the relationship of the two portfolios at the current time  $t = 0$ ?

5. If the stochastic equations

$$dx = pdt + \beta(x^2 + p) dV \quad (5.59)$$

$$dp = xdt - \gamma(x^3 + p^2) dW, \quad (5.60)$$

are Ito equations, then calculate the corresponding Stratonovich equations. Note that  $dW$  and  $dV$  are mutually independent Wiener noise processes.

6. The equations of motion for a damped harmonic oscillator are

$$dx = (p/m)dt, \quad (5.61)$$

$$dp = -kxdt - \gamma pdt. \quad (5.62)$$

Consider (i) a harmonic oscillator whose spring constant,  $k$ , is randomly fluctuating, and (ii) a harmonic oscillator whose damping rate,  $\gamma$ , is randomly fluctuating. Determine the Ito equations for the oscillator in both cases by first writing down the Stratonovich stochastic equations, and then transforming these to Ito equations.

## 6

### Numerical methods for Gaussian noise

#### 6.1 Euler's method

In Chapter 3 we described essentially all of the stochastic equations for which analytic solutions are presently known. Since this is a very small fraction of all possible stochastic equations, one must often solve these equations by simulating them using a computer (a process referred to as *numerical simulation*). To solve the stochastic equation

$$dx = f(x, t)dt + g(x, t)dW \quad (6.1)$$

in this way, we first choose an initial value for  $x$ . We can then approximate the differential equation by the equation

$$\Delta x = f(x, t)\Delta t + g(x, t)\Delta W, \quad (6.2)$$

where  $\Delta t$  is a small fixed value, and  $\Delta W$  is a Gaussian random variable with zero mean and variance equal to  $\Delta t$ . For each time-step  $\Delta t$ , we calculate  $\Delta x$  by using a random number generator to pick a value for  $\Delta W$ . We then add  $\Delta x$  to  $x$ , and repeat the process, continuing to increment  $x$  by the new value of  $\Delta x$  for each time-step.

The result of the numerical simulation described above is an approximation to a single sample path of  $x$ . To obtain an approximation to the probability density for  $x$  at some time  $T$ , we can repeat the simulation many times, each time performing the simulation up until time  $T$ , and each time using a different set of randomly chosen values for the  $\Delta W$ s. This generates many different sample paths for  $x$  between time 0 and  $T$ , and many different samples of the value of  $x$  at time  $T$ . We can then make a histogram of the values of  $x$ , and this is an approximation to the probability density for  $x(T)$ . We can also obtain an approximation to the mean and variance of  $x(T)$  simply by calculating the mean and variance of the samples we have generated.

The above method works just as well for a vector stochastic equation for a set of variables  $\mathbf{x} = (x_1, \dots, x_N)$ , driven by a vector of independent noise increments

$d\mathbf{W} = (dW_1, \dots, dW_M)$ :

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + G(\mathbf{x}, t)d\mathbf{W}, \quad (6.3)$$

where  $G$  is an  $N \times M$  matrix. As above the approximation to  $d\mathbf{x}$  is given by replacing  $dt$  with  $\Delta t$ , and  $d\mathbf{W}$  with  $\Delta\mathbf{W} = (\Delta W_1, \dots, \Delta W_M)$ .

This method for simulating stochastic differential equations is the stochastic equivalent of Euler's method [22] for deterministic differential equations. Below we will describe a more sophisticated method (Milstein's method) that is more accurate.

### 6.1.1 Generating Gaussian random variables

To realize the simulation described above, we need to generate Gaussian random variables with zero mean. The following method generates two independent zero mean Gaussian variables with variance  $\sigma^2 = 1$  [22]. We first take two random variables,  $x$  and  $y$ , that are uniformly distributed on the interval  $[0, 1]$  (All modern programming languages include inbuilt functions to generate such variables.) We then calculate

$$x' = 2x - 1, \quad (6.4)$$

$$y' = 2y - 1. \quad (6.5)$$

These new random variables are now uniformly distributed on the interval  $[-1, 1]$ . We now calculate

$$r = x'^2 + y'^2. \quad (6.6)$$

If  $r = 0$ , or  $r \geq 1$ , then we return to the first step and calculate new random variables  $x$  and  $y$ . If  $r \in (0, 1)$  then we calculate

$$g_1 = x' \sqrt{-2 \ln(r)/r}, \quad (6.7)$$

$$g_2 = y' \sqrt{-2 \ln(r)/r}. \quad (6.8)$$

The variables  $g_1$  and  $g_2$  are Gaussian with zero mean and unit variance, and mutually independent. If instead we want  $g_1$  and  $g_2$  to have variance  $c$ , then we simply multiply them by  $\sqrt{c}$ .

## 6.2 Checking the accuracy of a solution

The accuracy of the sample paths, means, and variances that we calculate using Euler's method will depend upon the size of the time-step  $\Delta t$ , and the values of  $x$ ,  $f$  and  $g$  throughout the evolution. The smaller the time-step with respect to these

values, then the more accurate the simulation. The approximate sample path given by the simulation converges to a true sample path in the limit as the time-step tends to zero. A simple way to check the accuracy of a simulation that uses a particular time-step is to perform the simulation again, but this time with the time-step halved. If the results of the two simulations differ by only a little, then the first simulation can be assumed to be accurate to approximately the difference between them. This is because we expect the result of halving the time-step a second time to change the result much less than it did the first time.

The process of halving the time-step,  $\Delta t$ , of a simulation deserves a little more attention. Note that a given sample path is generated by a given realization of the noise, and this is the set of the  $\Delta W_n$ s (chosen at random) for the simulation. Let us say that there are  $N$  time-steps in our simulation, and denote each of the  $N$  noise increments by  $\Delta W_n$ , where  $n = 0, \dots, N - 1$ . If we wish to halve the time-step, and generate an approximation to the *same* sample path, then we need to generate a set of  $2N$  Gaussian random numbers,  $\Delta \tilde{W}_m$ , that agree with the previous set of  $N$  random numbers  $\Delta W_n$ . What this means is that the sum of the first two stochastic increments  $\Delta \tilde{W}_0$  and  $\Delta \tilde{W}_1$  must equal the first stochastic increment  $\Delta W_0$ . This is because the total stochastic increment for the second simulation in the time-period  $\Delta t$ , is  $\Delta \tilde{W}_0 + \Delta \tilde{W}_1$ , and this must agree with the stochastic increment for the first simulation in the same time-step, in order for the two simulations to have the same noise realization. This must also be true for the second pair,  $\Delta \tilde{W}_2$  and  $\Delta \tilde{W}_3$ , etc. So we require

$$\Delta \tilde{W}_{2n} + \Delta \tilde{W}_{2n+1} = \Delta W_n, \quad n = 0, 1, \dots, N - 1. \quad (6.9)$$

An example of two approximations to the same sample path, one with half the time-step of the other, is shown in Figure 6.1. Fortunately it is very easy to generate a set of  $\Delta \tilde{W}_m$  for which Eq. (6.9) is true. All one has to do is generate  $N$  random numbers  $r_n$  with mean zero and variance  $\Delta t/2$ , and then set

$$\Delta \tilde{W}_{2n} = r_n \quad (6.10)$$

$$\Delta \tilde{W}_{2n+1} = \Delta W_n - r_n. \quad (6.11)$$

The above procedure allows one to perform two simulations of the same sample path for an SDE with different time-steps. If the difference between the final values of  $x$  for the two simulations are too large, then one can halve the time-step again and perform another simulation. One stops when the process of halving the time-step changes the final value of  $x$  by an amount that is considered to be small enough for the given application.

By repeatedly halving the time-step, one can also determine how rapidly the simulation converges to the true value of  $x(T)$ . The faster the convergence the

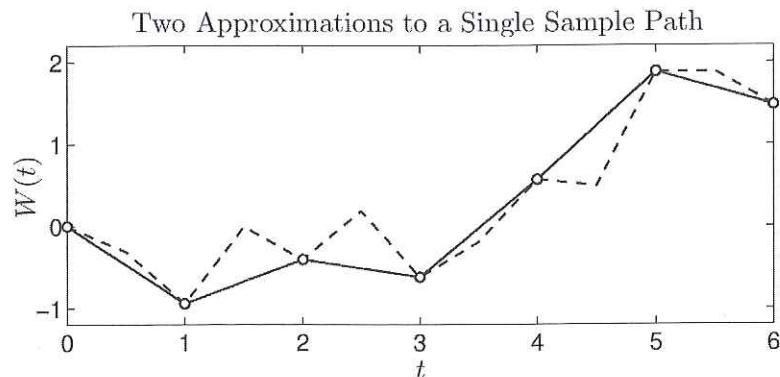


Figure 6.1. The solid line is a numerical approximation to a sample path of the Wiener process, with a time-step of  $\Delta t = 1$ . The dashed line is an approximation to the same sample path, but with half the time-step. For these approximations to correspond to the same sample path, the second must agree with the first at every multiple of  $\Delta t$ , denoted by the circles.

better, and different numerical methods have different rates of convergence. The simple Euler method that we described above has the slowest rate of convergence.

### 6.3 The accuracy of a numerical method

The accuracy of a numerical method for simulating an ordinary differential equation, stochastic or otherwise, is measured by how fast the numerical solution converges to the true solution as the time-step,  $\Delta t$ , is reduced. For a given differential equation, and a given sample path, the error in the final value of  $x(T)$  scales as some power of the time-step. The higher this power the faster the rate of convergence. The faster the convergence, the larger the time-step that can be used to achieve a given accuracy. The larger the time-step the *fewer* steps are required for the simulation, and (so long as each step is not hugely more complex for the faster method) the simulation can be performed in less time.

In general the error of the Euler method is only guaranteed to scale as  $\Delta t$  (but it sometimes does better – see Section 6.4.1 below). The accuracy of a numerical method is referred to as the *order* of the method, and for stochastic simulations, by convention, this order is quoted as one-half less than the power by which the error scales. The Euler method is thus said to be half-order. There is a perfectly sensible reason for this terminology. It comes from writing the approximate increment as a power series in  $\Delta t$ , which one imagines to agree with the true increment up to some particular power, and then disagree for higher powers. Because the  $\Delta W$  part of the increment is proportional to  $\sqrt{\Delta t}$ , one must use a power series in  $\sqrt{\Delta t}$ , rather than  $\Delta t$ . Thus if the error of the increment  $\Delta x$  is proportional to  $\Delta t$ , as in

Euler's method, then the increment is *accurate* to  $\sqrt{\Delta t}$  (half-order in  $\Delta t$ ), being the order of the first term in the power series. For a rigorous justification of these arguments, we refer the reader to the comprehensive text on numerical methods for SDEs by Kloeden and Platen [21].

In the next section we describe a simple numerical method that is accurate to first-order in  $\Delta t$ , and thus for which the error scales as  $(\Delta t)^{3/2}$ .

### 6.4 Milstien's method

The usefulness of Milstien's method comes from its simplicity, and the fact that it gives a significant improvement over Euler's method, since it is accurate to first-order in  $\Delta t$ . The Milstien method approximates the increment,  $dx$ , of the differential equation

$$dx = f(x, t)dt + g(x, t)dW \quad (6.12)$$

by

$$\Delta x = f \Delta t + g \Delta W + \frac{g}{2} \frac{\partial g}{\partial x} [(\Delta W)^2 - \Delta t]. \quad (6.13)$$

Here we have suppressed the arguments of  $f$  and  $g$  to avoid cluttering the notation.

#### 6.4.1 Vector equations with scalar noise

For the vector differential equation with a *single* (scalar) noise source,  $dW$ , given by

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + \mathbf{g}(\mathbf{x}, t)dW, \quad (6.14)$$

where  $\mathbf{x} = (x_1, \dots, x_N)^T$ ,  $\mathbf{f} = (f_1, \dots, f_N)^T$  and  $\mathbf{g} = (g_1, \dots, g_N)^T$ , Milstien's method is

$$\Delta x_i = f_i \Delta t + g_i \Delta W + \frac{1}{2} \sum_{j=1}^M g_j \frac{\partial g_i}{\partial x_j} [(\Delta W)^2 - \Delta t]. \quad (6.15)$$

Two special cases of this are as follows.

1. *The Ornstein–Uhlenbeck process with scalar noise.* The equation for this process is  $d\mathbf{x} = F\mathbf{x}dt + \mathbf{g}dW$ , where  $F$  is a constant matrix, and  $\mathbf{g}$  is a constant vector. As a result Milstien's method is just the same as Euler's method, being

$$\Delta \mathbf{x} = F\mathbf{x} \Delta t + \mathbf{g} \Delta W. \quad (6.16)$$

This means that for additive noise, Euler's method is a first-order method.

2. *The general linear stochastic equation with scalar noise.* For the general linear stochastic equation with a single noise source,  $\mathbf{dx} = F\mathbf{x}dt + G\mathbf{x}dW$ , where  $F$  and  $G$  are constant matrices, Milstien's method becomes

$$\Delta\mathbf{x} = F\mathbf{x}\Delta t + G\mathbf{x}\Delta W + \frac{1}{2}G^2\mathbf{x}[(\Delta W)^2 - \Delta t]. \quad (6.17)$$

#### 6.4.2 Vector equations with commutative noise

When there are *multiple* noise sources (that is, a vector of noise increments) Milstien's method is, in general, considerably more complex. Before we present this method in full, we consider a special case for which the method remains simple. A general vector stochastic differential equation is given by

$$\mathbf{dx} = \mathbf{f}(\mathbf{x}, t)dt + G(\mathbf{x}, t)\mathbf{dW}, \quad (6.18)$$

where  $\mathbf{dW}$  is a vector of mutually independent noise sources,

$$\mathbf{dW} = (dW_1, \dots, dW_M)^T, \quad (6.19)$$

and  $G$  is an  $N \times M$  matrix whose elements,  $G_{ij}$ , may be arbitrary functions of  $\mathbf{x}$  and  $t$ .

If the matrix  $G$  satisfies the set of relations

$$\sum_{m=1}^N G_{mj} \frac{\partial G_{ik}}{\partial x_m} = \sum_{m=1}^N G_{mk} \frac{\partial G_{ij}}{\partial x_m}, \quad \forall i, j, k, \quad (6.20)$$

then the noise is said to be *commutative*. A number of important special cases fall into this category, including *additive noise*, in which  $G$  is independent of  $\mathbf{x}$ , *diagonal noise* in which  $G$  is diagonal, and *separated linear noise*, in which

$$G_{ij} = g_{ij}(t)x_i, \quad (6.21)$$

where the  $g_{ij}(t)$  are arbitrary functions of time. This last case is referred to as “separated” because  $G_{ij}$  does not include  $x_j$ , for  $j \neq i$ .

When the noise is commutative, Milstien's method for solving Eq. (6.18) is

$$\begin{aligned} \Delta x_i &= f_i \Delta t + \sum_{j=1}^M G_{ij} \Delta W_j + \frac{1}{2} \sum_{j=1}^M \sum_{k=1}^M \left[ \sum_{m=1}^N G_{mj} \frac{\partial G_{ik}}{\partial x_m} \right] \Delta W_j \Delta W_k \\ &\quad - \frac{1}{2} \sum_{j=1}^M \left[ \sum_{m=1}^N G_{mj} \frac{\partial G_{ij}}{\partial x_m} \right] \Delta t. \end{aligned} \quad (6.22)$$

#### 6.4.3 General vector equations

For a completely general vector stochastic equation, the expression for Milstien's approximation to Eq. (6.18) involves a stochastic integral that cannot be written in terms of the discrete noise increments  $\Delta W_j$ . This expression is

$$\Delta x_i = f_i \Delta t + \sum_{j=1}^M G_{ij} \Delta W_j + \sum_{j=1}^M \sum_{k=1}^M \left[ \sum_{m=1}^N G_{mj} \frac{\partial G_{ik}}{\partial x_m} \right] \int_t^{t+\Delta t} \int_t^s dW_j(t') dW_k(s). \quad (6.23)$$

When  $j = k$ , the double integral does have a simple expression in terms of the discrete stochastic increments, being (see Section 3.8.3)

$$\int_t^{t+\Delta t} \int_t^s dW_j(t') dW_j(s) = \frac{1}{2} [(\Delta W_j)^2 - \Delta t]. \quad (6.24)$$

But this is not the case when  $j \neq k$ . Kloeden and Platen suggest a practical method for approximating the double integral for  $j \neq k$  [21], and this provides us with a numerical method for solving the vector stochastic equation. To present Kloeden and Platen's approximation we need a few definitions. We define  $a_{jm}$ ,  $b_{jm}$ , and  $c_{jm}$ , for each of the integer indices  $j$  and  $m$ , to be mutually independent Gaussian random variables with zero means and unit variances. We also define

$$\Upsilon_m = \frac{1}{12} - \frac{1}{2\pi^2} \sum_{n=1}^m \frac{1}{n^2}. \quad (6.25)$$

Finally, let us denote the double integral in Eq. (6.23) by

$$H_{jk} \equiv \int_t^{t+\Delta t} \int_t^s dW_j(t') dW_k(s). \quad (6.26)$$

With these definitions, an approximation to the double integral is given by

$$\begin{aligned} H_{jk}^{(m)} &= \left( \frac{\Delta W_j \Delta W_k}{2} + \sqrt{\frac{\Upsilon_m}{\Delta t}} [a_{jm} \Delta W_k - a_{km} \Delta W_j] \right) \\ &\quad + \frac{\Delta t}{2\pi} \sum_{n=1}^m \left( b_{jn} \left[ \sqrt{\frac{2}{\Delta t}} \Delta W_k + c_{kn} \right] - b_{kn} \left[ \sqrt{\frac{2}{\Delta t}} \Delta W_j + c_{jn} \right] \right). \end{aligned} \quad (6.27)$$

The larger the value of  $m$ , the better the approximation to  $H_{jk}$ . The average error of the approximation is given by

$$\langle (H_{jk}^{(m)} - H_{jk})^2 \rangle = \frac{(\Delta t)^2}{2\pi^2} \sum_{n=m+1}^{\infty} \frac{1}{n^2} \leq \frac{(\Delta t)^2}{2\pi^2 m}. \quad (6.28)$$

Here the bound is obtained by using the inequality

$$\sum_{n=m+1}^{\infty} \frac{1}{n^2} \leq \int_m^{\infty} \frac{1}{x^2} dx = \frac{1}{m}. \quad (6.29)$$

### 6.5 Runge–Kutter-like methods

A potential disadvantage of the Milstien method is that one must evaluate the first derivative of the function  $g(x, t)$  that multiplies the stochastic increment. For deterministic differential equations it is the Runge–Kutter family of methods that eliminate the need to evaluate such derivatives. Similar methods can be found for stochastic equations. Here we present a first-order method of this type that was obtained by Platen, building upon Milstien's method. We will refer to it as the Milstien–Platen method, and it is obtained from Milstien's method above by replacing the derivative of  $g$  with an approximation that is valid to first-order. For a stochastic equation containing only a single variable  $x$ , the first-order approximation to the term  $gg'$  that appears in Milstien's method is

$$g(x, t) \frac{\partial}{\partial x} g(x, t) \approx \frac{1}{\sqrt{\Delta t}} [g(q, t) - g(x, t)], \quad (6.30)$$

with

$$q = x + f \Delta t + g \sqrt{\Delta t}. \quad (6.31)$$

Substituting this into Milstien's method for a single variable, Eq. (6.13), we obtain the Milstien–Platen method for a single variable:

$$\Delta x = f \Delta t + g \Delta W + \frac{1}{2\sqrt{\Delta t}} [g(q, t) - g(x, t)] [(\Delta W)^2 - \Delta t]. \quad (6.32)$$

For a vector stochastic equation with scalar noise (Eq. (6.14)), the Milstien–Platen method is

$$\Delta x_i = f_i \Delta t + g_i \Delta W + \frac{1}{2\sqrt{\Delta t}} [g_i(\mathbf{q}, t) - g_i(\mathbf{x}, t)] [(\Delta W)^2 - \Delta t], \quad (6.33)$$

where the  $i$ th element of the vector  $\mathbf{q}$  is

$$q_i = x_i + f_i \Delta t + g_i \sqrt{\Delta t}. \quad (6.34)$$

For a general vector equation, the Milstien–Platen method is obtained by substituting into the Milstien method (Eq. (6.23)) the approximation

$$\sum_{m=1}^N G_{mj}(\mathbf{x}, t) \frac{\partial}{\partial x_m} G_{ik}(\mathbf{x}, t) \approx \frac{1}{\sqrt{\Delta t}} [G_{ij}(\mathbf{q}^{(k)}, t) - G_{ij}(\mathbf{x}, t)], \quad (6.35)$$

where the  $i$ th element of the vector  $\mathbf{q}^{(k)}$  is

$$q_i^{(k)} = x_i + f_i \Delta t + G_{ik} \sqrt{\Delta t}. \quad (6.36)$$

### 6.6 Implicit methods

Numerical methods for solving differential equations can suffer from instability. An instability is defined as an exponential (and thus rapid) increase in the error, and happens when a small error in one time-step causes a larger error in the next time-step, and so on. This can happen spontaneously at some point in time, even when the solution has been accurate up until that time. While this problem is fairly rare, it is more prevalent for differential equations that generate motion on two or more very different time-scales. (That is, the solution oscillates on a fast time-scale, and also changes on a much longer time-scale, where the longer time-scale is the one we are really interested in.) Differential equations with two or more disparate time-scales are referred to as being “stiff”.

If a particular method has a problem with instability, this can usually be fixed by using “implicit” methods. Implicit methods are just like the “explicit” methods we have discussed above, but they use the value of the variable or variables at the *end* of the time-step to evaluate the derivative, instead of the value(s) at the beginning. For example, instead of writing the Euler method for a single variable as

$$x(t + \Delta t) = x(t) + \Delta x(t) = x(t) + f[x(t), t] \Delta t + g[x(t), t] \Delta W, \quad (6.37)$$

we replace  $x(t)$  in the functions  $f$  and  $g$  with  $x(t + \Delta t)$ . This gives

$$x(t + \Delta t) = x(t) + f[x(t + \Delta t), t] \Delta t + g[x(t + \Delta t), t] \Delta W. \quad (6.38)$$

This method is called “implicit”, because it gives an “implicit” equation for  $x(t + \Delta t)$  in terms of  $x(t)$ . To actually calculate  $x(t + \Delta t)$  from  $x(t)$  we must solve Eq. (6.38) for  $x(t + \Delta t)$ . This is simple if  $f$  and  $g$  are linear functions of  $x$ , and if not, then we can always use a version of the Newton–Raphson method to solve the implicit equation at each time-step [22].

For a vector stochastic equation (Eq. (6.18)), it turns out that making the replacement  $\mathbf{x}(t) \rightarrow \mathbf{x}(t + \Delta t)$  in  $\Delta \mathbf{x}(t)$  (that is, in the functions  $\mathbf{f}$  and  $G$  in Eq. (6.18)) in *any* explicit method preserves the order of the method. Thus all the methods we have described above can be immediately turned into implicit methods by making this replacement.

### 6.7 Weak solutions

So far we have considered the speed at which a numerical method will converge to a sample path. Convergence to a sample path is referred to as *strong* convergence.

Similarly, methods that give strong convergence are called *strong methods*, and the sample paths they generate are called *strong solutions*. If one is not interested in the sample paths, but only in the properties of the probability density of  $x$  at some final time  $T$ , then one need only consider how fast these properties, such as the mean and variance, converge. These properties are determined by averaging over many sample paths, and for a given numerical method, often converge at a different rate than the individual sample paths. The convergence of the moments of the probability density for a stochastic process  $x(T)$  is referred to as *weak* convergence. Methods that provide this kind of convergence are called *weak methods*, and sample paths generated by these methods are called *weak solutions*. Euler's method, described above in Section 6.1, converges at first-order in  $\Delta t$  for the purposes of obtaining weak solutions, so long as the functions  $f(\mathbf{x}, t)$  and  $G(\mathbf{x}, t)$  in Eq. (6.3) have continuous fourth derivatives. Thus Euler's method is a half-order strong method, and a first-order weak method.

### 6.7.1 Second-order weak methods

We can, in fact, obtain second-order weak schemes that are no more complex, and in the general case less complex, than Milstien's strong scheme described above. For a stochastic equation for a single variable (Eq. (6.1)), a second-order method [21] is,

$$\begin{aligned}\Delta x &= f \Delta t + g \Delta W + \frac{1}{2} gg' [(\Delta W)^2 - \Delta t] \\ &\quad + \frac{1}{2} \left( ab' + ba' + \frac{1}{2} b^2 b'' \right) \Delta W \Delta t \\ &\quad + \frac{1}{2} \left( aa' + \frac{1}{2} b^2 a'' \right) (\Delta t)^2,\end{aligned}\quad (6.39)$$

where we have defined  $f' \equiv \partial f / \partial x$  and  $g' \equiv \partial g / \partial x$ .

For a vector stochastic equation (Eq. (6.18)), a second-order method is

$$\begin{aligned}\Delta x_i &= f_i \Delta t + Df_i(\Delta t)^2 + \sum_{j=1}^M G_{ij} \Delta W_j \\ &\quad + \frac{1}{2} \sum_{j=1}^M \left[ DG_{ij} + \sum_{k=1}^N G_{kj} \frac{\partial f_i}{\partial x_k} \right] \Delta W_j \Delta t \\ &\quad - \frac{1}{2} \sum_{j=1}^M \sum_{k=1}^M \left( \sum_{m=1}^N G_{mj} \frac{\partial G_{ik}}{\partial x_m} \right) (\Delta W_j \Delta W_k - \xi_{jk} \Delta t),\end{aligned}\quad (6.40)$$

where  $D$  is the differential operator

$$D = \frac{\partial}{\partial t} + \sum_{n=1}^N f_n \frac{\partial}{\partial x_n} + \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \sum_{l=1}^M G_{nl} G_{ml} \frac{\partial^2}{\partial x_n \partial x_m}. \quad (6.41)$$

The  $\xi_{jk}$  are given by

$$\xi_{ii} = 1, \quad (6.42)$$

$$\xi_{ij} = -\xi_{ji}, \quad \text{for } j < i, \quad (6.43)$$

and for  $j > i$  the  $\xi_{ij}$  are two-valued independent identical random variables with the probabilities

$$\text{Prob}(\xi_{ij} = \pm 1) = 1/2. \quad (6.44)$$

### Further reading

For all the classes of methods we have presented here, there are also higher-order versions, although for strong methods these become rather complex. The interested reader can find all these methods, along with full details of their derivations, in the comprehensive tour de force by Kloeden and Platen [21].