2

Stochastic Differential Equations and Related Inverse Problems

2.1 Concepts in Stochastic Calculus

As we have discussed in chapter 1, the deterministic mathematical formulation of solute transport through a porous medium introduces the dispersivity, which is a measure of the distance a solute tracer would travel when the mean velocity is normalized to be one. One would expect such a measure to be a mechanical property of the porous medium under consideration, but the evidence are there to show that dispersivity is dependent on the scale of the experiment for a given porous medium. One of the challenges in modelling the phenomena is to discard the Fickian assumptions, through which dispersivity is defined, and develop a mathematical discription containing the fluctuations associated with the mean velocity of a physical ensemble of solute particles. To this end, we require a sophisticated mathematical framework, and the theory of stochastic processes and differential equations is a natural mathematical setting. In this chapter we review some essential concepts in stochastic processes and stochastic differential equations in order to understand the stochastic calculus in a more applied context.

A deterministic variable expressed as a function of time uniquely determines the value of the variable at a given time. A stochastic variable Y, on the other hand, is one that does not have a unique value; it can have any one out of a set of values. We assign a unique label ω to each possible value of the stochastic variable, and set Ω to denote the set of all such values. When Y represents, for example the outcome of throwing dice, Ω may be a finite set of discrete numbers, and when Y is the instantaneous position of a fluid particle, it may be a continuous range of real numbers. If a particular value y_ω is observed for Y, this is called an event F. In fact, this is only the simplest prototype of an event; other possibilities might be that the value of Y is observed not to be y_ω (the complementary event), or that a value within a certain range of ω values is observed. The set of all possible events is denoted by F. Even though the outcome of a particular observation of Y is unpredictable, the probability of observing y_ω must be determined by a probability function $P(\omega)$. By using the standard methods of probability calculus, this implies that a probability $P(F)$ can also be assigned to compound events F e.g. by appropriate summation or integration over ω values. For this to work, F must satisfy the criteria that for any event F in its complement F^c must also belong to F, and that for any subset of F's the union of these must also belong to F. The explanation above of what it means to call Y a stochastic variable, is encapsulated in formal mathematical language by saying "Y is defined on a probability space (ω, F, P)" .

In describing physical systems, deterministic variables usually depend on additional parameters such as time. Similarly, a stochastic variable may depend on an additional

parameter t (for example, the probability may change with time, i.e. $P(y_\omega,t)$. The collection of stochastic variables, Y_t , is termed a stochastic process. The word 'process' suggests temporal development and is particularly appropriate when the parameter t has the meaning of time, but mathematically it is equally well used for any other parameter, usually assumed to be a real number in the interval $[0,\infty)$.

The label ω is often explicitly included in writing the notation $Y_t(\omega)$, for an individual value obtained from the set of Y-values at a fixed t. Conversely, we might keep ω fixed, and let t vary; a natural notation would be to write $Y_\omega(t)$. In physical terms, one may think of this as the set of values obtained from a single experiment to observe the time development of the stochastic variable Y. When the experiment is repeated, a different set of observations are obtained; those may be labelled by a different value of ω. Each such sequence of observed Y-values is called a realization (or sometimes a path) of the stochastic process, and from this perspective ω may be considered as labelling the realizations of the process. It is seen that it is somewhat arbitrary which of ω and t is considered to be a label, and which is an independent variable; this is sometimes expressed by writing the stochastic process as $Y(t,\omega)$.

In standard calculus, we deal with differentiable functions which are continuous except perhaps in certain locations of the domain under consideration. To understand the continuity of the functions better we make use of the definitions of the limits. We call a function f, a continuous function at the point $t = t_0$ if $\lim_{t\to t_0} f(t) = f(t_0)$ regardless of the direction t approaches t_0. A right-continuous function at t_0 has a limiting value only when t approaches t_0 from the right direction, i.e. t is larger than t_0 in the vicinity of t_0. We will denote this as

$$f(t+)=\lim_{t\downarrow t_0} f(t)=f(t_0) \ .$$

Similarly a left-continuous function at t_0 can be represented as

$$f(t-)=\lim_{t\uparrow t_0} f(t)=f(t_0) \ .$$

These statements imply that a continuous function is both right-continuous and left-continuous at a given point of t. Often we encounter functions having discontinuities; hence the need for the above definitions. To measure the size of a discontinuity, we define the term "jump" at any point t to be a discontinuity where the both $f(t+)$ and $f(t-)$ exist and the size of the jump be $\Delta f(t)=f(t+)-f(t-)$. The jumps are the discontinuities of the first kind and any other discontinuity is called a discontinuity of the second kind. Obviously a function can only have countable number of jumps in a given range. From the mean value theorem in calculus, it can be shown that we can differentiate a function in a given interval only if the function is either continuous or has a discontinuity of the second kind during the interval. Stochastic calculus is the calculus dealing with often non-differentiable functions having jumps without discontinuities of the second kind. One such example of a function is the Wiener process (Brownian motion). One realization of the standard Wiener process is given in Figure 2.1. These statements imply that a continuous function is both right-continuous and left-continuous at a given point of t. Often we encounter functions having

discontinuities; hence the need for the above definitions. To measure the size of a discontinuity, we define the term "jump" at any point t to be a discontinuity where the both $f(t+)$ and $f(t-)$ exist and the size of the jump be $\Delta f(t)=f(t+)-f(t-)$. The jumps are the discontinuities of the first kind and any other discontinuity is called a discontinuity of the second kind. Obviously a function can only have countable number of jumps in a given range. From the mean value theorem in calculus, it can be shown that we can differentiate a function in a given interval only if the function is either continuous or has a discontinuity of the second kind during the interval. Stochastic calculus is the calculus dealing with often non-differentiable functions having jumps without discontinuities of the second kind. One such example of a function is the Wiener process (Brownian motion). One realization of the standard Wiener process is given in Figure 2.1.

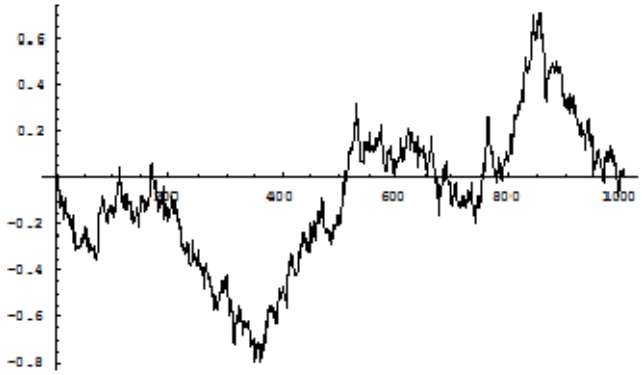

Figure 2.1. A realization of the Wiener process; this is a continuous but non-differentiable function.

The increments of the function shown in Figure 2.1 are irregular and a derivative cannot be defined according to the mean value theorem. This is because of the fact that the function changes erratically within small intervals, however small that interval may be. Therefore we have to devise new mathematical tools that would be useful in dealing with these irregular, non-differentiable functions.

Variation of a function f on $[a,b]$ is defined as

$$V_f([a,b])=\lim_{\delta_n\to 0}\sum_{i=1}^{n}\left|f(t_i^n)-f(t_{i-1}^n)\right| \tag{2.1.1}$$

where $\delta_n=\max_{1\le i\le n}(t_i-t_{i-1})$.

If $V_f([a,b])$ is finite such as in continuous differentiable functions then f is called a function of finite variation on $[a,b]$. Variation of a function is a measure of the total change in the function value within the interval considered. An important result (Theorem 1.7 Klebaner (1998)) is that a function of finite variation can only have a countable number of jumps. Furthermore, if f *is* a continuous function, f' exists and $\int|f'(t)|dt<\infty$ then f is a function

of finite variation. This implies that a function of finite variation on [*a,b*] is differentiable on [*a,b*], and a corollary is that a function of infinite variation is non-differentiable. Another mathematical construct that plays a major role in stochastic calculus is the quadratic variation. In stochastic calculus, the quadratic variation of a function *f* over the interval [*0,t*] is given by

$$[f](t)=\lim_{\delta_n\to 0}\sum_{i=1}^{n}(g(t_i^n)-g(t_{i-1}^n))^2 \quad , \tag{2.1.2}$$

where the limit is taken over the partitions:

$$0=t_0^n<t_1^n<\ldots<t_n^n=t \, ,$$

with $\delta_n=\max_{1\le i\le n}(t_i^n-t_{i-1}^n)\to 0$.

It can be proved that the quadratic variation of a continuous function with finite variation is zero. However, the functions having zero quadratic variation may have infinite variation such as zero energy processes (Klebaner, 1998). If a function or process has a finite positive quadratic variation within an interval, then its variation is infinite, and therefore the function is continuous but not differentiable.

Variation and quadratic variation of a function are very important tools in the development of stochastic calculus, even though we do not use quadratic variation in standard calculus.

We also define quadratic covariation of functions *f* and *g* on [0,*t*] by extending equation (2.1.2):

$$[f,g](t)=\lim_{\delta_n\to 0}\sum_{i=0}^{n-1}(f(t_{i+1}^n)-f(t_i^n))(g(t_{i+1}^n)-g(t_i^n)) \tag{2.1.3}$$

when the limit is taken over partitions $\{t_i^n\}$ of [0,*t*] with $\delta_n=\max_{1\le i\le n}(t_{i+1}^n-t_i^n)\to 0$. It can be shown that if both the functions are continuous and one is of finite variation, the quadratic covariation is zero.

Quadratic covariation of two functions, *f* and *g,* has the following properties:

1. Polarization identity

Polarization identity expresses the quadratic covariation, $[f,g](t)$, in terms of quadratic variation of individual functions.

$$[f,g](t)=\frac{1}{2}([f+g,f+g](t)-[f,f](t)-[g,g](t)) \tag{2.1.4}$$

2. Symmetry

$$[f,g](t) = [g,f](t) \, . \tag{2.1.5}$$

3. Linearity

Using polarization identity and symmetry one can show that covariation is linear for any constants *a* and *b*,

$$[af+bg,h](t)=a[f,h](t)+b[g,h](t)\ . \tag{2.1.6}$$

Quadratic variation of a function $[f](t)$ and covariation $[f,g](t)$ are measures of change in the functional values over a given range $[0,t]$.

In many situations where stochastic processes are involved, we would like to know the limiting values of useful random variables, i.e. whether they approach a some sort of a "steady state" or asymptotic behaviour. It is natural to define the steady state of random variable within a probabilistic context. Therefore, in stochastic processes, we define the convergence of random variables using the following four different criteria:

1. Almost sure convergence

Random variables $\{X_n\}$ converges to $\{X\}$ with probability one:

$$P(\{\omega\in\Omega:\lim_{n\to\infty}|X_n(\omega)-X(\omega)|=0\})=1\ .$$

2. Mean-square convergence

$\{X_n\}$ converges to $\{X\}$ such a way that $E(X_n^2)<\infty$ for $n = 1,2,...,n$, $E(X)<\infty$ and

$$\lim_{n\to\infty}E(|X_n-X|^2)=0\ .$$

3. Convergence in probability

$\{X_n\}$ converges to $\{X\}$ with zero probability of having a difference between the two processes:

$$\lim_{n\to\infty}P(\{\omega\in\Omega;|X_n(\omega)-X(\omega)|\geq\varepsilon)=0\ ,\text{ for all }\ \varepsilon>0.$$

Convergence in probability is called stochastic convergence as well.

Note that we adopt the notation of E(,) or E[,] to denote the expected value (mean value) of a stochastic variable. In physical literature, this is denoted by "< , >".

4. Convergence in distribution

Distribution function of $\{X_n\}$ converges to that of $\{X\}$ at any point of continuity of the limiting distribution (i.e. the distribution function of $\{X\}$).

These four criteria add another dimension to our discussion of the asymptotic behaviour of a process. These arguments can be extended in comparing stochastic processes with each other.

Unlike in deterministic variables where any asymptotic behaviour can clearly be identified either graphically or numerically, stochastic variables do require adherence to one of the convergence criteria mentioned above which are called the "criteria for strong convergence". There are weakly converging stochastic processes and we do not discuss the weak convergence criteria as they are not relevant to the development of the material in this book.

In standard calculus we have continuous functions with discontinuities at finitely many points and we integrate them using the definition of Riemann integral of a function $f(t)$ over the interval $[a,b]$:

$$\int_a^b f(t)\,dt = \lim_{\delta\to 0} \sum_{i=1}^{n} f(\xi_i^n)\,(t_i^n - t_{i-1}^n), \tag{2.1.7}$$

where t_i^n's represents partitions of the interval,

$$a = t_0^n < t_1^n < t_2^n \ldots < t_n^n = b\ ,$$

$$\delta = \max_{1\le i\le n}(t_i^n - t_{i-1}^n), \quad \text{and} \qquad t_{i-1}^n \le \xi_i^n \le t_i^n \quad .$$

Riemann integral is used extensively in standard calculus where continuous functions are the main concern. The integral converges regardless of the chosen ξ_i^n within $[\,t_{i-1}^n, t_i^n\,]$.

A generalization of Riemann integral is Stieltjes integral which is defined as the integral of $f(t)$ with respect to a monotone function $g(t)$ over the interval $[a,b]$:

$$\int_a^b f(t)\,dg(t) = \lim_{\delta\to 0} \sum_{i=1}^{n} f(\xi_i)(g(t_i^n) - g(t_{i-1}^n)) \tag{2.1.8}$$

with the same definitions as above for δ and t_i^n's. It can be shown that for the Stieltjes integral to exist for any continuous function $f(t)$, $g(t)$ must be a function with finite variation on $[a,b]$. This means that if $g(t)$ has infinite variation on $[a,b]$ then for such a function, integration has to be defined differently. This is the case in the integration of the continuous stochastic processes, therefore, can not be integrated using Stieltjes integral. Before we discuss alternative forms of integration that can be applied to the functions of positive quadratic variation, i.e. the functions of infinite variation, we introduce a fundamentally important stochastic process, the Wiener process and its properties.

2.2 Wiener Process

The botanist Robert Brown, first observed that pollen grains suspended in liquid, undergo irregular motion. Centuries later, it was realised that the physical explanation of this is that the pollen grain is continually bombarded by molecules of the liquid travelling with different speeds in different directions. Over a time scale that is large compared with the intervals between molecular impacts, these will average out and no net force is exerted on the grain. However, this will not happen over a small time interval; and if the mass of the grain is small enough to undergo appreciable displacement in the small time interval as the result of molecular impacts, an observable erratic motion results. The crucial point to notice in the present context is that while the impacts and therefore the individual

displacements suffered by the grain can be considered independent at different times, the actual position of the grain can only change continuously.

In the physical Brownian motion, there are small but nevertheless finite intervals between the impulses of molecules colliding with the pollen grain. Consequently, the path that the grain follows, consists of a sequence of straight segments forming an irregular but continuous line – a so-called random walk. Each straight segment can be considered an increment of the momentary position of the grain.

The mathematical idealisation of the Brownian motion let the interval between increments approach zero. The resulting process – called the Wiener process due to N. Wiener – is difficult to conceptualise: for example, consideration shows that the resulting position is everywhere continuous, but nowhere differentiable. This means that while the particle has a position at any moment, and since this position is changing it is moving, yet no velocity can be defined. Nevertheless as discussed by Stroock and Varadhan (1979) a consistent mathematical description is obtained by defining the position as a stochastic process $B(t,\omega)$ with the following properties that are suggested as a mathematical model for the Brownian motion- a Wiener process:

P1: $B(0,\omega) = 0$, i.e. choose the position of the particle at the arbitrarily chosen initial time $t = 0$ as the coordinate origin;

P2: $B(t,\omega)$ has independent increments, i.e. $B(t_1,\omega)$, $\{B(t_2,\omega) - B(t_1,\omega)\},\ldots, \{B(t_k,\omega) - B(t_{k-1},\omega)\}$ are independent for all $0 \leq t_1 < t_2 \ldots < t_k$;

P3: $\{B(t_{i+1},\omega) - B(t_i,\omega)\}$ is normally distributed with mean 0 and variance $(t_{i+1} - t_i)$;

P4: The stochastic variation of $B(t,\omega)$ at fixed time t is determined by a Gaussian probability;

P5: The Gaussian has a zero mean, $E[B(t,\omega)] = 0$ for all values of t;

P6: $B(t,\omega)$ are continuous functions of t for $t \geq 0$;

P7: The covariance of Brownian motion is determined by a correlation between the values of $B(t,\omega)$ at times t_i and t_j (for fixed ω), given by

$$E\left[B(t_i,\omega)\ B(t_j,\omega)\right] = \min(t_i,t_j). \tag{2.2.1}$$

When applied to $t_i = t_j = t$, P7 reduces to the statement that

$$Var\left[B(t,w)\right] = t, \tag{2.2.2}$$

where '*Var*' means statistical variance. For the Brownian motion this can be interpreted as the statement that the radius within which the particle can be found increases proportional to time.

Because the Wiener process is defined by the independence of its increments, it is for some purposes convenient to reformulate the variance of a Wiener process in terms of the variance of the increments:

From P3, for $t_i < t_j$:

$$Var[B(t_j,\omega) - B(t_i,\omega)] = t_j - t_i \tag{2.2.3}$$

Bearing in mind that the statistical definition of the variance of a quantity X reduces to the expectation value expression $Var[X]=E[X^2]-(E[X])^2$ and that the expectation value or mean of a Wiener process is zero, we can rewrite this as,

$$E[\{B(t_2,\omega) - B(t_1,\omega)\}^2] = Var[B(t_2,\omega) - B(t_1,\omega)], \quad \text{i.e.} \quad E[\Delta B \cdot \Delta B] = \Delta t\,, \tag{2.2.4}$$

where Δt is defined to mean the time increment for a fixed realization ω.

The connection between the two formulations is established by similarly rewriting equation (2.2.3) and then applying equation (2.2.1):

$$\begin{aligned} Var[B(t_1,\omega) - B(t_2,\omega)] &= E[\{B(t_1,\omega) - B(t_2,\omega)\}^2] \\ &= E[B^2(t_1,\omega) + B^2(t_1,\omega) - 2B(t_1,\omega)B(t_2,\omega)] \\ &= t_1 + t_2 - 2\min(t_1,t_2) \\ &= t_1 - t_2 \text{ for } t_1 > t_2\,. \end{aligned}$$

2.3 Further Properties of Wiener Process and their Relationships

Consider a stochastic process $X(t,\omega)$ having a stationary joint probability distribution and $E(X(t,\omega)) = 0$, i.e. the mean value of the process is zero. The Fourier transform of $Var(X(t,\omega))$ can be written as,

$$S(\lambda,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} Var(X(\tau,\omega)\, e^{-i\lambda\tau} d\tau \quad . \tag{2.3.1}$$

$S(\lambda,\omega)$ is called the spectral density of the process $X(t,\omega)$ and is also a function of angular frequency λ . The inverse of the Fourier transform is given by

$$Var(X(\tau,\omega)) = \int_{-\infty}^{\infty} S(\lambda,\omega)\, e^{i\lambda\tau} d\lambda\,. \tag{2.3.2}$$

Therefore variance of $X(0,\omega)$ is the area under a graph of spectral density $S(\lambda,\omega)$ against λ :

$$Var(X(0,\omega)) = E(X^2(0,\omega)) \quad \text{because} \quad E(X(t,\omega))=0\,.$$

Spectral density $S(\lambda,\omega)$ is considered as the "average power" per unit frequency at λ which gives rise to the variance of $X(t,\omega)$ at $\tau = 0$. If the average power is a constant, the power is distributed uniformly across the frequency spectrum, such as the case for white light, then $X(t,\omega)$ is called white noise. White noise is often used to model independent random disturbances in engineering systems, and the increments of Wiener process have the same characteristics as white noise. Therefore white noise $(\zeta(t))$ is defined as

$$\zeta(t) = \frac{dB(t)}{dt} \quad , \quad \text{and} \quad dB(t) = \zeta(t)dt \, . \tag{2.3.3}$$

We will use this relationship to formulate stochastic differential equations.

As shown before, the relationships among the properties mentioned above can be derived starting from P1 to P7. For example, let us evaluate the covariance of Wiener processes, $B(t_i, \omega)$ and $B(t_j, \omega)$:

$$Cov(B(t_i, \omega)\, B(t_j, \omega)) = E(B(t_i, \omega)\, B(t_j, \omega)) \, . \tag{2.3.4}$$

Assuming $t_i < t_j$, we can express:

$$B(t_j, \omega) = B(t_i, \omega) + B(t_j, \omega) - B(t_i, \omega) \, . \tag{2.3.5}$$

Therefore,

$$\begin{aligned} E(B(t_i, \omega)\, B(t_j, \omega)) &= E(B(t_i, \omega)(B(t_i, \omega) + B(t_j, \omega) - B(t_i, \omega)) \\ &= E(B^2(t_i, \omega) + B(t_i, \omega)B(t_j, \omega) - B^2(t_i, \omega)) \\ &= E(B^2(t_i, \omega) + B(t_j, \omega)(B(t_i, \omega) - B(t_j, \omega))) \\ &= E(B^2(t_i, \omega)) + E(B(t_j, \omega)(B(t_i, \omega) - B(t_j, \omega))) \end{aligned} \tag{2.3.6}$$

From P2, $B(t_j, \omega)$ and $(B(t_i, \omega) - B(t_j, \omega))$ are independent processes and therefore we can write ,

$$E(B(t_j, \omega)(B(t_i, \omega) - B(t_j, \omega))) = E(B(t_i, \omega))E(B(t_i, \omega) - B(t_j, \omega)) \quad . \tag{2.3.7}$$

According to P3 and P5, $E(B(t_j, \omega)) = 0$ and $E(B(t_i, \omega) - B(t_j, \omega)) = 0$.

Therefore, from equation (2.3.7)

$$E(B(t_j, \omega)B(t_i, \omega) - B(t_j, \omega))) = 0 \, .$$

This leads equation (2.3.6) to $E(B(t_i, \omega)B(t_j, \omega)) = E(B^2(t_i, \omega))$,

$$\text{And} \quad E(B^2(t_i, \omega)) = E((B(t_i, \omega)) - 0)^2) \, . \tag{2.3.8}$$

From P3, $\{B(t_i, \omega) - B(0, \omega)\}$ is normally distributed with a variance $(t_i - 0)$, and equation (2.3.8) becomes, $E(B^2(t_i, \omega)) = t_i$, and , therefore, $Cov(B(t_i, \omega)B(t_j, \omega)) = t_i$.

Using a similar approach it can be shown that if $t_i > t_j$,

$$Cov(B(t_i, \omega)B(t_j, \omega)) = t_j \, . \tag{2.3.9}$$

This leads to P7: $E(B(t_i, \omega)B(t_j, \omega)) = \min(t_i, t_j)$.

The above derivations show the relatedness of the variance of an independent increment, $Var\{B(t_1, \omega) - B(t_2, \omega)\}$ to the properties of Wiener process given by P1 to P7. The fact that

$\{B(t_{i+1},\omega)-B(t_i,\omega)\}$ is a Gaussian random variable with zero mean and $\{t_{i+1}-t_i\}$ variance can be used to construct Wiener process paths on computer. If we divide the time interval $[0,t]$ into n equidistant parts having length Δt, and at the end of each segment we can randomly generate a Brownian increment using the Normal distribution with mean 0 and variance Δt. This increment is simply added to the value of Wiener process at the point considered and move on to the next point. When we repeat this procedure starting from $t=\Delta t$ to t=t and taking the fact that $B(0,\omega)=0$ into account, we can generate a realization of Wiener process. We can expect these Wiener process realizations to have properties quite distinct from other continuous functions of t. We will briefly discuss some important characteristics of Wiener process realizations next as these results enable us to utilise this very useful stochastic process effectively.

Some useful characteristics of Wiener process realizations $B(t,\omega)$ are

1. $B(t,\omega)$ is a continuous , nondifferentiable function of t.

2. The quadratic variation of $B(t,\omega)$, $[B(t,\omega),B(t,\omega)](t)$ over $[0,t]$ is t.

Using the definition of covariation of functions,

$$\begin{aligned}[B(t,\omega),B(t,\omega)](t) &= [B(t,\omega),B(t,\omega)]([0,t]) \\ &= \lim_{\delta_n\to 0}\sum_{i=1}^{n}[B(t_i^n)-B(t_{i-1}^n)]^2 \quad .\end{aligned} \tag{2.3.10}$$

where $\delta_n=\max(t_{i+1}^n-t_i^n)$ and $\{t_i^n\}_{i=1}^n$ is a partition of $[0\ ,t]$, as $n\to\infty$, $\delta_n\to o$.

Taking the expectation of the summation,

$$E(\sum(B(t_i^n)-B(t_{i-1}^n))^2)=\sum(E((B(t_i^n)-B(t_{i-1}^n))^2)). \tag{2.3.11}$$

$E((B(t_i^n)-B(t_{i-1}^n))^2)$ is the variance of an independent increment $\{B(t_i^n)-B(t_{i-1}^n)\}$.

As seen before, $Var[B(t_i^n)-B(t_{i-1}^n)]=(t_i^n-t_{i-1}^n)$.

Therefore,

$$\begin{aligned}E\left(\sum(B(t_i^n)-(B(t_{i-1}^n))^2\right) &= \sum Var[B(t_i^n)-B(t_{i-1}^n)] \\ &= \sum_{i=1}^{n}(t_i^n-t_{i-1}^n)=t-0=t.\end{aligned} \tag{2.3.12}$$

Let us take the variance of $\sum(B(t_i^n)-B(t_{i-1}^n))^2$:

$$Var\left(\sum(B(t_i^n)-B(t_{i-1}^n))^2\right)=\sum 3(t_i^n-t_{i-1}^n)^2\le 3\ \max(t_i^n-t_{i-1}^n)\ t=3t\delta_n\text{, and}$$

as $n\to\infty,\delta_n\to 0,\sum Var(B(t_i^n)-B(t_{i-1}^n))^2\to 0$.

Summarizing the results,

$$E(\sum (B(t_i^n) - B(t_{i-1}^n))^2) = t,$$

and

$$Var(\sum (B(t_i^n) - B(t_{i-1}^n))^2) \to 0 \quad \text{as} \quad n \to \infty .$$

This implies that, according to the monotone convergence theories that $\sum (B(t_i^n) - B(t_{i-1}^n))^2 \to t$ almost surely as $n \to \infty$.

Therefore, the quadratic variation of Brownian motion $B(t,\omega)$ is *t*:

$$[B(t,\omega), B(t,\omega)](t) = t . \tag{2.3.13}$$

Omitting *t* and ω, $[B,B](t) = t$.

3. Wiener process $(B(t,\omega))$ is a martingale.

A stochastic process, $\{X(t)\}$ is a martingale, when the future expected value of $\{X(t)\}$ is equal to {X (*t*)}. In mathematical notation, $E(X(t+s)|F_t) = X(t)$ with converging almost surely, F_t *is* the information about $\{X(t)\}$ up to time *t*. We do not give the proof of these martingale characteristics of Brownian motion here but it is easy to show that $E(B(t+s) \mid F_t) = B(t)$.

It can also be shown that $\{B(t,\omega)^2 - t\}$ and $\{\exp(\alpha B(t,\omega) - \frac{\alpha^2}{2}t)\}$ are also martingales. These martingales can be used to characterize the Wiener process as well and more details can be found in Klebaner (1998).

4. Wiener process has Markov property

Markov property simply states that the future of a process depends only on the present state. In other words, a stochastic process having Markov property does not "remember" the past and the present state contains all the information required to drive the process into the future states.

This can be expressed as

$$P(X(t+s) \le y \mid F_t) = P(X(t+s) \le y \mid X(t)) , \tag{2.3.14}$$

Converging almost surely.

From the very definition of increments of the Wiener process for the discretized intervals of [0,t], $\{B(t_{i+1}^n) - B(t_i^n)\}$, the Wiener process increment behaves independently to its immediate predecessor $\{B(t_i^n) - B(t_{i-1}^n)\}$.

In other words $\{B(t_{i+1}^n) - B(t_i^n)\}$ does not remember the behaviour of $\{B(t_i^n) - B(t_{i-1}^n)\}$ and only element common to both increments is $B(t_i^n)$.

One can now see intuitively why Wiener process should behave as a Markov process. This can be expressed as

$$P(B(t_i + s) \le x_i \mid \{B(t_i), B(t_{i-1})...0)\}) = P(B(t_i + s) \le x_i \mid B(t_i)) , \tag{2.3.15}$$

which is another way of expressing the previous equation (2.3.14).

5. Generalized form of Wiener process

The Wiener process as defined above is sometimes called the standard Wiener process, to distinguish it from that obtained by the following generalization:

$$E[\ B(t_i,\omega)\ B(t_j,\omega)\ ] = \int_0^{\min(t_i,t_j)} q(\tau)\, d\tau\ .$$

The integral kernel $q(\tau)$ is called the correlation function and determines the correlation between stochastic process values at different times. The standard Wiener process is the simple case that $q(\tau)$=1 , i.e. full correlation over any time interval; the generalised Wiener process includes, for example, the case that q decreases, and there is progressively less correlation between the values in a given realization as the time interval between them increases.

2.4 Stochastic Integration

At this point of our discussion, we need to define the integration of stochastic process with respect to the Wiener process $(B(t,\omega))$ so that we understand the conditions under which this integral exists and what kind of processes can be integrated using this integral. The Stieltjes integral can not used to integrate the functions of infinite variation, and therefore, there is a need to define the integrals for the stochastic process such as the Wiener process. There are two choices available: Ito definition of integration and Stratanovich integration. These two definitions produce entirely different integral stochastic process.

The Ito definition is popular among mathematicians and physicists tend to use the Stratanovich integral. The Ito integral has the martingale property among many other useful technical properties (Keizw, 1987), and in addition, the Stratanovich integrals can be reduced to Ito integrals (Klebaner, 1998). In this monograph, we confine ourselves to Ito definition of integration:

$$I[X](\omega) = \int_S^T X(t,\omega)\, dB(t,\omega)\ .$$

$I[X](\omega)$ implies that the integration of $X(t,\omega)$ is along a realization ω and with respect to the Wiener process (a.k.a Brownian motion) which is a function of t. $I[X](\omega)$ is also a stochastic process in its own right and have properties originating from the definition of the integral. It is natural to expect $I[X](\omega)$ to be equal to $c(B(t,\omega) - B(s,\omega))$ when $X(t,\omega)$ is a

constant *c*. If $X(t)$ is a deterministic process, which can be expressed as a sequence of constants over small intervals, we can define Ito integral as follows:

$$\begin{aligned} I[X] &= \int_S^T X(t)\,dB(t) \\ &= \sum_{i=0}^{n-1} c_i((B(t_{i+1}) - B(t_i))) \end{aligned} \quad , \tag{2.4.1}$$

where $X(t) = \begin{cases} c_0, & t = S \\ c_i, & t_i < t \le t_{i+1} \end{cases}$ $i = 0, \cdots, n-1$.

The time interval $[S,T]$ has been discretized into *n* intervals : $S = t_0 < t_1 < \cdots < t_n = T$.

Using the property of independent increments of Brownian motion, we can show that the mean of $I[X](\omega)$ is zero and,

$$Variance = Var(I[X]) = \sum_{i=0}^{n-1} c_i^2 (t_{i+1} - t_i) . \tag{2.4.2}$$

It turns out that if $X(t,\omega)$ is a continuous stochastic process and its future values are solely dependent on the information of this process only up to *t,* Ito integral $I[X](\omega)$ exists. The future states on a stochastic process, $X(t,\omega)$, is only dependent on F_t then it is called an adapted process. A left-continuous adapted process $X(t,\omega)$ is defined as a predictable process and it satisfies the following condition: $\int_0^T X^2(t,\omega)\,dt < \infty$ with almost surely convergence.

As we are only concerned about continuous processes driven by the past events, we limit our discussion of predictable processes. Reader may want to refer to ∅ksendal (1998) and Klebaner (1998) for more rigorous discussion of these concepts.

We can now define Ito integral $I[X](\omega)$ similarly to equation (2.4.1) if $X(t,\omega)$ is a continuous and adapted process then $I[X](\omega)$ can be defined as

$$\sum_{i=0}^{n-1} X(t_i^n,\omega)(B(t_{i+1}^n,\omega) - B(t_i^n,\omega)) \ , \tag{2.4.3}$$

and this sum converges in probability.

Dropping ω for convenience and adhering to the same discretization of interval [*S, T*] as in equation (2.4.1),

$$I[X] = \int_S^T X(t)dB(t) = \sum_{i=0}^{n-1} X(t_i^n)\,(B(t_{i+1}^n) - B(t_i^n)) \ . \tag{2.4.4}$$

Equation (2.4.4) expresses an approximation of $\int_S^T X(t)\,dB(t)$ based on the convergence in probability. We take equation (2.4.3) as the definition of Ito integral for the purpose of this

book. As stated earlier $I[X](\omega)$ is a stochastic process and it has the following properties (see, for example, Øksendal (1998) for more details):

1. Linearity

If $X(t)$ and $Y(t)$ are predictable processes and α and β as some constants, then

$$I[\alpha X + \beta Y](\omega) = \alpha\, I[X](\omega) + \beta\, I[Y](\omega)\,. \tag{2.4.5}$$

2. Zero mean Property

$$E(I[X](\omega)) = 0\,. \tag{2.4.6}$$

3. Isometry Property

$$E[(\int_S^T X(t)\,dB(t))^2] = \int_S^T E(X^2(t))\,dt\,. \tag{2.4.7}$$

The isometry property says that the expected value of the square of Ito integral is the integral with respect to t of the expectation of the square of the process $X(t)$.

Since $E[(\int_S^T X(t)\,dB(t))] = 0$ from zero mean property, we can express the left hand side of equation (2.4.7) as

$$\begin{aligned} &E((\int_S^T X(t)dB(t))^2 - E(\int_S^T X(t)dB(t))) \\ &= E[\int_S^T X(t)dB(t) - E(\int_S^T X(t)dB(t))]^2 = Var\left(\int_S^T X(t)dB(t)\right) \end{aligned}\,. \tag{2.4.8}$$

Therefore the variance of Ito integral process is $\int_s^T E(X^2(t))dt$ and this result will be useful to us in understanding the behaviour of Ito integral process. We say that Ito integral is square integrable. According to Fubuni's Theorem, which states that, for a stochastic process $X(t)$, with continuous realizations,

$$E(\int_s^T X(t)dt) = \int_s^T E(X(t))dt\,, \tag{2.4.9}$$

and

$$E(\int_s^T X^2(t)dt) = \int_s^T E(X^2(t))dt\,. \tag{2.4.10}$$

4. Ito integral is a martingale

It can be shown that $E(I[X(t)]\,|\,F_t) = I[X(t)]$. Strictly speaking $X(t)$ should satisfy $\int_S^T X^2(t)dt < \infty$ and $\int_S^T E(X^2(t))dt < \infty$ for martingale property to be true. Therefore, Ito integrals are square integrable martingales.

5. Ito integral of a deterministic function $X(t)$ is a Guassian process with zero mean and covariance function,

$$Cov(I[X(t)], I[X(t+t_0)]) = \int_0^t X^2(s)ds\ ,\ \ t_0 \geq 0. \tag{2.4.11}$$

$I\ [X\ (t\ )\ ]$ is a square integrable martingale.

6. Quadratic variation of Ito integral,

$$[I[X], I[X]](t) = \int_0^T X^2(t)dt\ . \tag{2.4.12}$$

We see that Ito integral has a positive quadratic variation making it a process with infinite variation i.e., it is a nondifferentiable continuous function of t.

7. Quadratic covariation of Ito integral with respect to processes $X_1(t)$ and $X_2(t)$ is given by

$$[I[X_1], I[X_2]](t) = \int_0^T X_1(t)X_2(t)dt\ . \tag{2.4.13}$$

Armed with these properties we can proceed to discuss the machinery of stochastic calculus such as stochastic chain rule, which is also known as Ito formula.

2.5 Stochastic Chain Rule (Ito Formula)

As we have seen previously, the quadratic variations of Brownian motion, $[B(t, \omega), B(t, \omega)](t)$, is the limit in probability over the interval $[0,t\ ]$:

$$[B(t,\omega), B(t,\omega)](t) = \lim_{\delta_n \to 0} \sum_{i=0}^{n-1} (B(t_{i+1}^n) - B(t_i^n))^2\ , \tag{2.5.1}$$

$\delta_n = \max(t_{i+1}^n - t_i^n) \to 0$. Using the differential notation, $\Delta B = B(t_{i+1}^n) - B(t_i^n)$, and summation as an integral ,

$$[B(t,\omega), B(t,\omega)](t) = \int_0^t (dB(s))^2\ . \tag{2.5.2}$$

We have shown that $[B,B]\ (t) = t$, and therefore, $\int_0^t (dB(s))^2 = t$.

For our convenience and also to make the notation similar to the one in standard differential calculus, we denote

$$\int_0^t (dB(s))^2 = t \tag{2.5.3}$$

$$\text{as} \qquad (dB(t))^2 = dt\ . \tag{2.5.4}$$

This equation does not have a meaning outside the integral equation (2.5.3) and should not be interpreted in any other way.

Similarly for any other continuous function $g(t)$,

$$g(t)(dB(t))^2 = g(B(t))dt\ , \tag{2.5.5}$$

which means,

$$g(t)(dB(t))^2 = g(B(t))dt\ . \tag{2.5.6}$$

This equation is an expression of the approximation, converging in probability, of

$$\lim_{\delta_n \to 0} \sum_{i=0}^{n-1} g(t_i^n)(B(t_{i+1}^n) - B(t_i^n))^2 = \int_0^t g(B(s))ds\ . \tag{2.5.7}$$

As the quadratic variation of a continuous and differentiable function is zero,

$$[t,t]\ (t) = 0. \tag{2.5.8}$$

This equation in integral notation,

$$\int_0^t (dt)^2 = 0\ ,$$

and in differential notation,

$$(dt)^2 = 0\ . \tag{2.5.9}$$

Similarly, quadratic covariation of t (a continuous and differentiable function) and Brownian notion,

$$[t,B](t) = 0\ . \tag{2.5.10}$$

This relationship can be proved by expressing quadratic covariation as

$$[t,B](t) = \lim_{\delta_n \to o} \sum_{i=0}^{n-1} (t_{i+1}^n - t_i^n)(B(t_{i+1}^n) - B(t_i^n))$$

$$\delta_n = \max(t_{i+1}^n - t_i^n)\ ,$$

$$\begin{aligned} [t,B](t) &\le \delta_n \sum_{i=0}^{n-1} (B(t_{i+1}^n) - B(t_i^n)) \\ &\le \delta_n B(t) \end{aligned}\ .$$

Therefore as $n \to \infty$, $\delta_n \to 0$ (because t is a function of finite variation),

$$[t,B](t) \to 0 \quad \text{as} \quad n \to \infty\ .$$

Hence, $[t,B](t) = 0$ and in integral notation, $\int_0^t dt\ dB = 0$.

This can be written in differential notation,

$$dt.dB = 0\ . \tag{2.5.11}$$

Therefore, we can summarize the following rules in differential notation as follows,

$$dt.dt = 0\,;\quad dt.dB = 0\,;\quad dB.dt = 0,\quad \text{and}\quad dB.dB = dt\,. \tag{2.5.12}$$

In order to come to grips with the interpretation of the differential properties of dB_t, it is useful to consider the chain rule of differentiation. This will also lead us to formulas that are often more useful in calculating Ito integrals than the basic definition as the limit of a sum. Consider first the case in ordinary calculus of a function $g(x,t)$, where x is also a function of t. We can write the change in g as t changes, as follows:

$$\Delta g = \frac{\partial g(t,x)}{\partial t}\Delta t + \frac{\partial g(t,x)}{\partial x}\Delta x + \tfrac{1}{2}\frac{\partial^2 g(t,x)}{\partial x^2}(\Delta x)^2 + \ldots$$

From this, an expression for dg/dt is obtained by taking the limit $\Delta t \to 0$ of the ratio $(\Delta g/\Delta t)$.

Since $\Delta x = (dx/dt)\,\Delta t$, when $\Delta t \to 0$ the 2nd derivative term shown is of order $(\Delta t)^2$ and falls away together with all higher derivatives, and the well-known chain rule formula for the total derivative (dg/dt) is obtained. However, if , instead of x , we have a Wiener process B_t , we get

$$\Delta g = \frac{\partial g(t,B_t)}{\partial t}\Delta t + \frac{\partial g(t,B_t)}{\partial x}\Delta B_t + \tfrac{1}{2}\frac{\partial^2 g(t,B_t)}{\partial x^2}(\Delta B_t \cdot \Delta B_t) + \ldots$$

If the expectation value of this expression over all realizations is taken, the above shows that the second derivative term is now only of order Δt and cannot be ignored. Since this holds for the expectation value, for consistency we also cannot neglect the term if the limit $\Delta t \to 0$ is taken without considering the expectation value. Unlike the case of ordinary calculus where all expressions containing products of differentials higher than 1 is neglected, in Ito calculus we therefore have different rules.

Recall that in standard calculus chain rule is applied to composite functions.

For example, if $Y=f(t)$ then $g(Y)$ is a function of Y.

Then $\dfrac{dg}{dt} = \dfrac{dg}{dY}.\dfrac{dY}{dt}$

In differential notation,

$$dg = \frac{dg}{dY}.df\,.$$

By integrating

$$g(f(t)) = g(0) + \int_0^t g'(f(t))df\,.$$

Suppose say $f(t) = B(t)$ (Brownian motion) and $g(x)$ is twice continuously differentiable function. Then by using stochastic Taylor series expansion,

$$g(B(t)) = g(0) + \int_0^t g'(B(s))dB(s) + \frac{1}{2}\int_0^t g''(B(s))ds\,. \tag{2.5.13}$$

Comparing equation (2.5.13) and the corresponding stochastic chain rule, we can see that the second derivative term of the Taylor series plays a significant role in changing the chain rule in the standard calculus to the stochastic one.

For example, let $g(x) = e^x$

$$\text{Therefore, } e^{B(t)} = e^{(0)} + \int_0^t e^{B(s)} dB(s) + \frac{1}{2}\int_0^t e^{B(s)} ds \,. \tag{2.5.14}$$

In differential notation (which is only a convention),

$$d(e^{B(t)}) = e^{B(t)} dB(t) + \frac{1}{2} e^{B(t)} dt \,. \tag{2.5.15}$$

As an another example, let $g(x) = x^2$.

Therefore, from the chain rule

$$(B(t))^2 = (B(0))^2 + 2\int_s^t B(s) dB(s) + \frac{1}{2}\int_0^t 2ds \,,$$

$$\int_0^t B(s) dB(s) = \frac{1}{2}(B(t))^2 - \frac{1}{2} t \,. \tag{2.5.16}$$

This is quite a different result from the standard integration. In differential convention,

$$B(t)\, dB(t) = \frac{1}{2} d((B(t))^2) - \frac{1}{2} dt \,. \tag{2.5.17}$$

In other words, the stochastic process $\int_0^t B(s) dB(s)$ can be calculated by evaluating $\{\frac{1}{2}(B(t))^2 - \frac{1}{2} t\}$. We will show how this process behaves using computer simulations in section 2.6.

We can write Ito integral as

$$Y(t) = \int_0^t \sigma(s) dB(s) \,. \tag{2.5.18}$$

Then we can add a "drift term" to the "diffusion term" given by equation (2.5.18):

$$Y(t) = Y(0) + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dB(s) \,. \tag{2.5.19}$$

We recall that $\sigma(s)$ should be a predictable process and is subjected to the condition $\int_0^T \sigma^2(t) dt < \infty$ converging almost surely. $\mu(t)$ is, on the other hand, an adapted continuous process of finite variation. In equation (2.5.19) $\int_0^t \sigma(s) dB(s)$ represents the diffusion part of the process and $\int_0^t \mu(s) ds$ does not contain the noise; therefore it represents

the drifting of the process. $Y(t)$ is called an Ito process and in differential notation we can write,

$$dY(t) = \mu(t)dt + \sigma(t)dB(t)\,. \tag{2.5.20}$$

Equation (2.5.20) can be quite useful in practical applications where the main driving force is perturbed by an irregular noise. A particle moving through a porous medium is such an example. In this case, advection gives rise to the drift term and hydrodynamic dispersion and microdiffusion give rise to the "diffusive" term. In the population dynamics, the diffusive term is a direct result of noise in the proportionality constant. Therefore it is important to study Ito process further in order to apply it in modeling situations. $\mu(t)$ is called the drift coefficient and $\sigma(t)$ the diffusion coefficient and they can depend on $Y(t)$ and/or $B(t)$. For example, we can write in pervious result (equation (2.5.17)),

$$d(B(t)^2) = dt + 2B(t)dB(t)\,. \tag{2.5.21}$$

This is an Ito process with the drift coefficient of 1 and the diffusion coefficient of $2B(t)$. Quadratic variation of Ito process on $[0,T]$

$$Y(t) = Y(0) + \int_0^t \mu(s)ds + \int_0^t \sigma(s)dB(s). \tag{2.5.22}$$

is given by

$$[Y,Y](t) = \int_0^t \sigma^2(s)ds \quad . \tag{2.5.23}$$

This can be deduced from the fact that $\int_0^t \mu(s)ds$ is a continuous function with finite variation and using quadratic variation of Ito integral. In differential notation,

$$\begin{aligned}(dY(t))^2 &= dY(t).dY(dt),\\ &= \mu^2(t)(dt)^2 + 2\mu\,\sigma dtdB + \sigma^2(dB)^2,\\ &= \sigma^2(t)dt.\end{aligned} \tag{2.5.24}$$

The chain rule given in equation (2.5.12) gives us a way to compute the behaviour of a function of Brownian motion. It is also useful to know the chain rule to compute a function of a given Ito process. Suppose an Ito process is given by a general form,

$$dX(t) = \mu dt + \sigma dB(t)\,, \tag{2.5.25}$$

where μ is the drift coefficient and σ is the diffusion coefficient and let $g(t, x)$ is a twice differentiable continuous function. Let $Y(t) = g(t, X(t))$. Here $Y(t)$ is a function of t and Ito process $X(t)$, and is also a stochastic process. $Y(t)$ can also be expressed as an Ito process. Then Ito formula states,

$$dY(t) = \frac{dg}{dt}(t,X(t))dt + \frac{\partial g}{\partial x}(t,X(t))dX(t) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X(t)).(dX(t))^2. \tag{2.5.26}$$

$$\text{where,}\quad (dX(t))^2 = d(X(t)).d(X(t))\,, \tag{2.5.27}$$

and is evaluated according to the rules given by equation (2.5.12).

As an example, consider the Ito process

$$dX(t) = dt + 2B(t)dB(t) \tag{2.5.28}$$

where $\mu = 1$ and $\sigma = 2B(t)$.

Assume $g(t,x) = x^2$, therefore,

$$\frac{\partial g}{\partial t} = 0\,;\ \frac{\partial g}{\partial x} = 2x\ \ ;\ \frac{\partial^2 g}{\partial^2 x} = 2\,. \tag{2.5.29}$$

Substituting to Ito formula above ,

$$\begin{aligned} dg &= 2X(t)dX(t) + dX(t).dX(t), \\ &= 2(dt + 2B(t)dB(t))B(t) + 4B^2(t)dt. \end{aligned} \tag{2.5.30}$$

$$dX^2(t) = (2X(t) + 4B^2(t))dt + \ 4X(t)B(t)dB(t). \tag{2.5.31}$$

As seen above $dX^2(t)$ is also an Ito process with $u = 2X(t) + 4B^2(t)$ (drift coefficient), a function of $X(t)$ and B (t), and v = $4X(t)B(t)$ (diffusion coefficient) , also a function of $X(t)$ and $B(t)$.

Substituting $X(t) = B^2(t)$ to equation (2.5.31),

$$\begin{aligned} d(B^4(t)) \ &= 2(B^2(t) + 2B^2(t)dt + 4(B^2(t))B(t)dB(t) \\ &= 6B^2(t)dt + 4B^3(t)dB(t) \end{aligned}\ . \tag{2.5.32}$$

We can derive this from chain rule for a function of B(t) as well.

Let $g(x) = x^4$, and from Ito formula:

$$\begin{aligned} dg &= g'(B(t))dB(t) + \frac{1}{2}g''(t)dt \\ &= 4B^3(t)dB(t) + \frac{1}{2}4.3.B^2(t)dt, \end{aligned} \tag{2.5.33}$$

$$d(B^4(t)) = 6B^2(t)dt + 4B^3(t)dB(t)\ . \tag{2.5.34}$$

This is the same Ito process as in equation (2.5.32). Let us consider another example which will be useful. Consider the function $g(x) = \ln x$ and the Ito process

$$dX(t) = \frac{1}{2}X(t) + X(t)dB(t)\,. \tag{2.5.35}$$

For this Ito process $\mu = \frac{1}{2}X(t)$ and $\sigma = X(t)$.

From the Ito formula (equation (2.5.26)),

$$\begin{aligned} d(\ln X(t)) &= \frac{1}{X(t)}dX(t) + \frac{1}{2}\left(-\frac{1}{X^2(t)}\right)(X^2(t)dt), \\ &= \frac{1}{X(t)}\left(\frac{1}{2}X(t)dt + X(t)dB(t)\right) - \frac{1}{2}dt, \\ &= \frac{1}{2}dt + dB(t) - \frac{1}{2}dt, \\ &= dB(t). \end{aligned} \tag{2.5.36}$$

By convention, the above stochastic differential is given by the following integral equation:

$$\ln X(t) = \ln X(0) + \int_0^t dB(t), \tag{2.5.37}$$

$$\ln\left[\frac{X(t)}{X(0)}\right] = B(t), \tag{2.5.38}$$

$$X(t) = X(0)e^{B(t)}.$$

We can show that $X(t) = X(0)e^{B(t)}$ satisfies $dX(t) = \frac{1}{2}X(t)dt + X(t)dB(t)$. In other words $X(t) = X(0)e^{B(t)}$ is a "solution" to the stochastic differential $dX(t) = \frac{1}{2}X(t)dt + X(t)dB(t)$.

This idea of having a solution to a stochastic differential is similar to having a solution to differential equations in standard calculus.

Suppose $X_1(t)$ and $X_2(t)$ are Ito processes given by the following differentials:

$$dX_1(t) = \mu_1(t)dt + \sigma_1(t)dB(t), \tag{2.5.39}$$

$$dX_2(t) = \mu_2(t)dt + \sigma_2(t)dB(t). \tag{2.5.40}$$

Quadratic covariation is given by

$$\begin{aligned} d[X_1, X_2] &= dX_1(t).dX_2(t), \\ &= \mu_1\mu_2(dt)^2 + \mu_1\,\sigma_2 dt.dB(t) + \mu_2\,\sigma_1 dt.dB(t) + \sigma_1\sigma_2(dB(t))^2. \end{aligned}$$

And $(dt)^2 = dt.dB(t) = 0$.

$$\begin{aligned} d[X_1, X_2] &= \sigma_1(t)\sigma_2(t)(dB(t))^2 \\ &= \sigma_1(t)\sigma_2(t)dt \end{aligned}. \tag{2.5.41}$$

The stochastic product rule is given by,

$$\begin{aligned} &X_1(t)X_2(t) - X_1(0)X_2(0) \\ &= \int_0^t X_1(s)dX_2(s) + \int_0^t X_2(s)dX_1(s) + [X_1, X_2](t) \end{aligned}. \tag{2.5.42}$$

If at least one of X_1 and X_2 is a continuous function with finite variation, then $[X_1, X_2](t) = 0$ and equation (2.5.42) reduces to the integration by parts formula in the standard calculus.

Stochastic product rule can be expressed in differential form:

$$\begin{aligned} &d(X_1(t)\,X_2(t)) \\ &= X_1(t)dX_2(t) + X_2(t)dX_1(t) + \sigma_1(t)\sigma_2(t)dt. \end{aligned} \tag{2.5.43}$$

As an example, consider $Y(t) = t\,B(t)$,

$$Y(t) = X_1(t)X_2(t)\,,$$

where $X_1(t) = t$, a continuous function with finite variation and $\sigma_1 = 0$, and $X_2(t) = B(t)$, Brownian motion with infinite variation and $\sigma_2 = 1$.

From the product rule,

$$d(Y(t)) = tdB(t) + B(t)dt + (0)(1)dt\,,$$

$$d(tB(t)) = tdB(t) + B(t)dt \tag{2.5.44}$$

This is the same result we obtain if we use the standard product rule. The reason for this is that quadratic covariation of a continuous function and a function with infinite variation is zero as we have mentioned previously.

Suppose $dX_1(t) = tdB(t) + B(t)dt$, and

$dX_2(t) = \frac{1}{2}X_2(t)dt + X_2(t)dB(t)$, where

$\mu_1(t) = B(t)$; $\sigma_1(t) = t$; $\sigma_2(t) = X_2(t)$; and $\mu_2(t) = \frac{1}{2}X_2$.

From the product rule,

$$\begin{aligned} d(X_1(t)X_2(t)) &= X_1(t)dX_2(t) + X_2(t)dX_1(t) + \sigma_1\sigma_2 dt, \\ &= X_1(t)dX_2(t) + X_2(t)dX_1(t) + tX_2(t)dt. \end{aligned} \tag{2.5.45}$$

By substitution,

$$\begin{aligned} &d(X_1(t)X_2(t)) = X_1(t)(\frac{1}{2}X_2dt + X_2(t)dB(t)) + X_2(t)(tdB(t) + B(t)dt) + tX_2(t)dt, \\ &= \left(\frac{1}{2}X_1(t)X_2(t) + X_2(t)B(t) + tX_2(t)\right)dt + (X_1(t)X_2(t) + tX_2(t))dB(t). \end{aligned} \tag{2.5.46}$$

This is again an Ito process.

$$\begin{aligned} d(X_1(t)X_2(t)) &= \left(\frac{1}{2}X_1(t)X_2(t) + tX_2(t) + X_2(t)B(t)\right)dt + \left(X_1(t)+t\right)X_2(t)dB(t), \\ &= X_2(t)\left(\frac{1}{2}X_1(t) + t + B(t)\right)dt + X_2(t)\left(X_1(t)+t\right)dB(t)). \end{aligned} \tag{2.5.47}$$

As an integral equation,

$$X_1(t)X_2(t) - X_1(0)X_2(0) = \int_0^t X_2(t)(\frac{1}{2}X_1(t) + t + B(t))dt + \int_0^t X_2(t)(X_1(t)+t)dB(t).$$

If $g(x_1, x_2)$ is a continuous and twice differentiable function of x_1 and x_2, and we are given Ito processes of the forms, $dX_1(t) = \mu_1 dt + \sigma_1 dB(t)$ and $dX_2(t) = \mu_2 dt + \sigma_2 dB(t)$.

Then $g(X_1(t), X_2(t))$ is also an Ito process and given by the following differential form:

$$\begin{aligned} dg(X_1(t), X_2(t)) &= \frac{\partial g(X_1)}{\partial x_1}dX_1(t) + \frac{\partial g(X_2)}{\partial x_2}dX_2(t) + \frac{1}{2}\frac{\partial^2 g(X_1)}{\partial^2 x_1}(dX_1(t))^2 \\ &+ \frac{1}{2}\frac{\partial^2 g(X_2)}{\partial^2 x_2}(dX_2(t))^2 + \frac{\partial^2 g(X_1X_2)}{\partial x_1 x_2}dX_1(t).dX_2(t). \end{aligned} \tag{2.5.48}$$

Using quadratic variation and covariation of Ito processes,

$$(dX_1(t))^2 = dX_1(t)\,.\,dX_1(t) = \sigma_1^2 dt\,,$$

$$(dX_2(t))^2 = dX_2(t)\,.\,dX_2(t) = \sigma_2^2 dt\,, \text{ and}$$

$$dX_1(t)\,.\,dX_2(t) = \sigma_1\sigma_2 dt\,.$$

These can be considered as a generalization of the rules on differentials given by equation (2.5.12). We use this generalized Ito formula for a function of two Ito processes in the following example.

We will express the stochastic process $X(t) = 2 + t + e^{B(t)}$ as an Ito process having the standard form, $dX(t) = \mu dt + \sigma dB(t)$.

We can consider

$$X(t) = g(t, B(t)) = 2 + t + e^{B(t)}\,. \tag{2.5.49}$$

Therefore, $g(t, y) = 2 + t + e^y$, where

$$X_1(t) = t\,,$$

$$X_2(t) = y = B(t)\,.$$

These equations give, $dX_1 = dt$ and $dX_2 = dB(t)$, where $\mu_1 = 1$; $\sigma_1 = 0$; $\mu_2 = 0$; and $\sigma_2 = 1$.

Using equation (2.5.48),

$$dg = (1)dt + e^{B(t)}dB(t) + \frac{1}{2}(0)(dB(t))^2 + \frac{1}{2}e^{B(t)}(dB(t))^2 + (0)dt.dB(t)$$
$$= dt + e^{B(t)}dB(t) + \frac{1}{2}e^{B(t)}dt.$$

Using $(dB(t))^2 = dt$,

$$dg = dX(t) = \left(1 + \frac{1}{2}e^{B(t)}\right)dt + e^{B(t)}dB(t).$$

From a previous example, $d(e^{B(t)}) = e^{B(t)}dB(t) + \frac{1}{2}e^{B(t)}dt$.

Therefore $dX(t) = dt + (\frac{1}{2}e^{B(t)}dt + e^{B(t)}dB(t)) = dt + d(e^{B(t)})$.

From the integral notation,

$$X(t) = X(0) + \int_0^t dt + \int_0^t d(e^{B(t)}),$$

$$X(t) = (0) + t + e^{B(t)} - 1,$$

$$X(t) = (X(0) - 1) + t + e^{B(t)}.$$

Comparing with

$$X(t) = 2 + t + e^{B(t)},$$

$$X(0) - 1 = 2,$$
$$X(0) = 3.$$

X(t)= constant + t + $e^{B(t)}$ can be considered as a solution process to the stochastic differential,

$$dX(t) = (1 + \frac{1}{2}e^{B(t)})dt + e^{B(t)}dB(t).$$

As we can see in the above solution, the solution process contains the characteristics of both the drift and diffusion phenomena. In this case, diffusion phenomenon dominates as t increases because of the expected value of the exponential of Brownian motion increases at a faster rate in general. If we examine the drift term of the stochastic differential above, we see that the drift term is also affected by the Brownian motion, so the final solution is always a result of complex interactions between the drift term and the diffusion term.

We now to discuss a population dynamics example equipped with the knowledge of Ito process and formula:

$$\frac{dx(t)}{dt} = \alpha(t)x(t)\,. \tag{2.5.50}$$

If the coefficient α (f) is "noisy", we can express if as follows:

$$\alpha(t) = r(t) + \sigma(t)W_t\,,$$

where W_t = white noise, then

$$dX(t) = (r(t)dt + \sigma(t)dB(t))\,.\,X(t)\,,$$

and Brownian motion increments $dB(t) = W_t dt$.

Therefore ,

$$dX(t) = \left(r(t)dt + \sigma(t)W_t dB(t)\right)X(t)\,,$$

$$dX(t) = r(t)X(t)dt + \sigma(t)X(t)d(t)\,. \tag{2.5.51}$$

As seen from the above equation (2.5.51), $X(t)$ is an Ito process.

Consider the case with $r(t)$= r , a constant and $\sigma(t) = \sigma$, a constant then the process $X_1(t)$ can be written in the differential form:

$$dX(t) = rX(t)dt + \sigma X(t)dB(t)\,.$$

Assume $g(x)$=$ln\ x$,

Then using the Ito formula,

$$dg(x(t)) = \frac{\partial(\ln x)}{\partial x}dX(t) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(dX(t))^2\,,$$

$$\begin{aligned} d(\ln(X(t)) &= \frac{dX(t)}{X(t)} + \frac{1}{2}\left(\frac{-1}{X(t)^2}\right)(\sigma^2), \\ &= \frac{dX(t)}{X(t)} - \frac{\sigma^2}{2}dt, \\ &= \frac{1}{X(t)}(rX(t)dt + \sigma X(t)dB(t)) - \frac{\sigma^2}{2}dt. \end{aligned}$$

$$\begin{aligned} d(\ln(X(t)) &= rdt + \sigma dB(t) - \frac{\sigma^2}{2}dt, \\ &= \left(r - \frac{\sigma^2}{2}\right)dt + \sigma dB(t). \end{aligned}$$

Converting back to the integral form,

$$\ln(X(t)) = \ln(X(0)) + \int_0^t \left(r - \frac{\sigma^2}{2} \right) dt + \int_0^t \sigma dB(t) \,,$$

$$\ln\left(\frac{X(t)}{X(0)} \right) = \left(r - \frac{\sigma^2}{2} \right) t + \sigma B(t) \,,$$

$$X(t) = X(0) \exp\left(\sigma B(t)\right) \exp\left((r - \frac{\sigma^2}{2}) t \right) . \tag{2.5.52}$$

$X(t)$ process, therefore, satisfies the Ito process,

$$dX(t) = rX(t)dt + \sigma X(t) dB(t) \,,$$

and equation (2.5.52) can be considered as a solution to the stochastic differential equation. As discussed earlier, this solution significantly different from its deterministic counterpart.

This section we have reviewed the essentials of stochastic calculus and presented the results which could be useful in developing models and solving stochastic differential equations. While analytical expressions are quite helpful to understand stochastic processes, computer simulation provides us with an intuitive "feel" for the simulated phenomena. Sometimes it is revealing to simulate a number of realizations of a process and visualize them on computers to understand the behaviours of the process.

2.6 Computer Simulation of Brownian Motion and Ito Processes

In the previous section, we have introduced Brownian motion (the Wiener process) as a stationary, continuous stochastic process with independent increments. This process is a unique one to model the irregular noise such as Gaussian white noise in systems, and once such a process is incorporated in differential equations, the process of obtaining solutions involve stochastic calculus. Only a limited number of stochastic differential equations have analytical solutions and some of these equations are given by Kloeden and Platen (1992). In many instances we have to resort to numerical methods. We illustrate the behaviour of the Wiener process and Ito processes through computer simulations so that reader can appreciate the variable nature of individual realizations.

For the numerical implementation, it is most convenient to use the variance specification of the Wiener process $B(t)$. The time span of the simulation, [0,1] is discretised into small equal time increments *delt*, and the corresponding independent Wiener increments selected randomly from a normal distribution with zero mean and variance, *delt*.

Figure 2.2 shows the Wiener process increments as a single stochastic process. Since Gaussian white noise is the derivative of Wiener process and the time interval is a constant, Figure 2.2 depicts a realization of an approximation of white noise process.

The Wiener process is very irregular (Figure 2.1), and the only discernible pattern is that as time progresses, the position tends to wander away from the starting position at the origin. In other words, if the statistical variance over realisations for a fixed time is evaluated, this

increases gradually – a property referred to as time varying variance. The use of the Wiener process in a modelling situation to represent the noise in the system should be carefully thought through. If the noise can be represented as white noise, then Wiener process enters into the equation because of the relationship between the white noise and the Wiener process. It is also important to realize that the Ito integral is a stochastic process dependent on the Wiener process. This is analogous to integration in standard calculus because an indefinite integral is a function of the independent, deterministic variable.

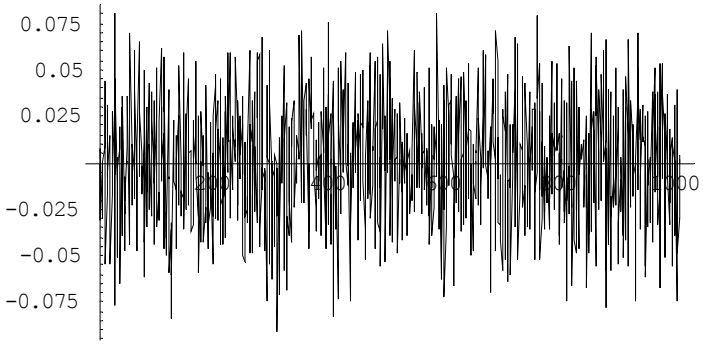

Figure 2.2. A realization of the Wiener process increment.

Given the Wiener process realization depicted in Figure 2.1, we compute the Ito integral of Wiener process, $\int_0^t B(t,\omega)dB$.

As we have previously seen, this integral can be evaluated by using the following stochastic relationship converging in probability; and it is shown in Figure 2.4.

$$\int_0^t B(s,\omega)dB(s,\omega) = \frac{1}{2}B^2(t,\omega) - \frac{1}{2}t \ .$$

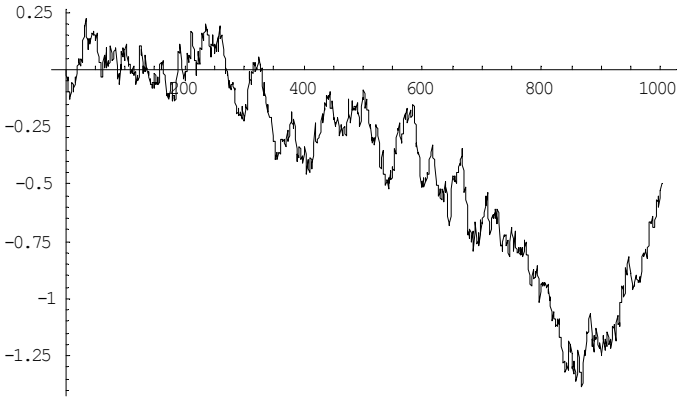

Figure 2.3. The realization of the Wiener process used in the calculation of the Ito Integral shown in Figure 2.4.

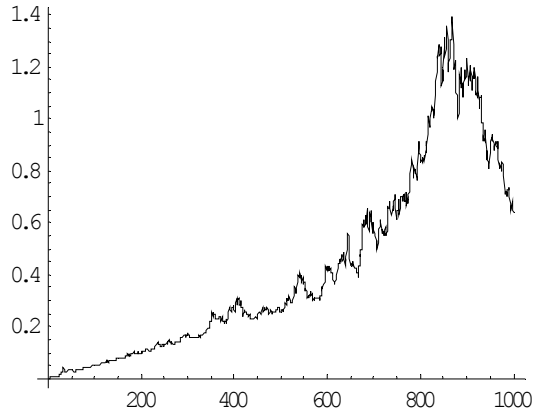

Figure 2.4. A realization of $\int_0^t B(t,\omega)dB$.

Let us consider the following Ito process which we have derived in section 2.5. In differential notation,

$$d(B^4(t))=6B^2(t)dt+4B^3(t)dB(t),$$

which means,

$$B^4(t)=B^4(0)+\int_0^t 6B^2(t)dt+\int_0^t 4B^3(t)dB(t), \text{ and}$$

$$B^4(t)=\int_0^t 6B^2(t)dt+\int_0^t 4B^3(t)dB(t). \tag{2.6.1}$$

The Ito process given in equation (2.6.1) is simulated in Figure 2.6 for the Wiener realization depicted in Figure 2.5.

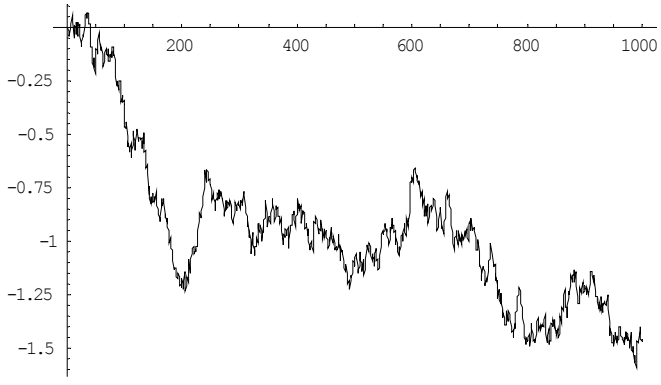

Figure 2.5. Wiener realization used in evaluating the Ito process $B^4(t)$.

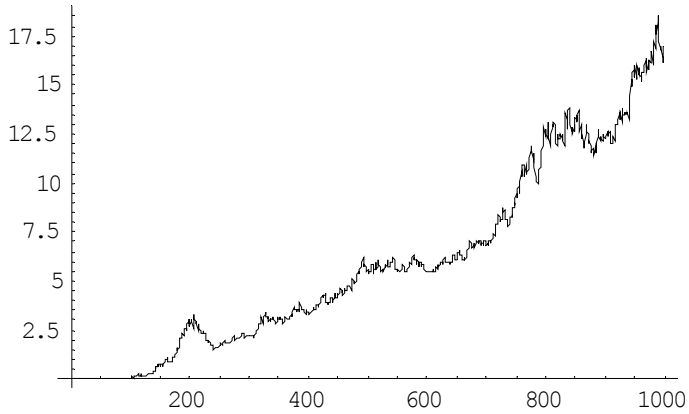

Figure 2.6. Ito process $B^4(t) = \int_0^t 6B^2(t)dt + \int_0^t 4B^3(t)dB(t)$.

Even for a decreasing and erratic Wiener process, the Ito process $\left\{\int_0^t 6B^2(t)dt + \int_0^t 4B^3(t)dB(t)\right\}$ in general has a smoother realization which has an overall growth in positive direction. The effect of Ito integration tends to smoothen the erratic behaviour of Wiener process. We have evaluated the above Ito process for 3 different realizations of the standard Wiener process, and they are shown in Figure 2.7.

As seen in Figure 2.7, individual realizations of the Ito process $\left\{\int_0^t 6B^2(t)dt + \int_0^t 4B^3(t)dB(t)\right\}$ are distinct from each other; and they show the complexity in stochastic integration as opposed to integration in standard calculus.

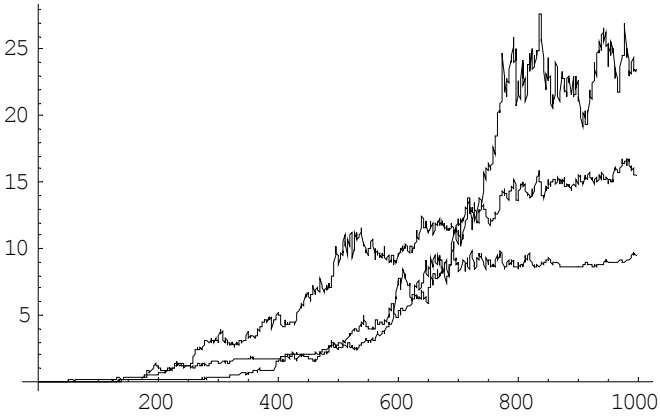

Figure 2.7. Three realizations of $\left\{\int_0^t 6B^2(t)dt + \int_0^t 4B^3(t)dB(t)\right\}$.

2.7 Solving Stochastic Differential Equation

Let us consider an ordinary differential equation which relates the derivative of the dependent variable ($y(t)$) to the independent variable (t) through a function, $\phi(y(t),t)$, with the initial condition $y(0) = y_0$:

$$\frac{dy}{dt} = \phi(y,t), \tag{2.7.1}$$

$$\text{and} \quad dy = \phi(y,t)dt. \tag{2.7.2}$$

In many natural systems, this rate of change can be influenced by random noise caused by a combination of factors, which could be difficult to model. As a model of this random fluctuations, white noise ($\xi(t)$) is a suitable candidate. Therefore we can write, in general, the increments of the noise process as $\sigma(y,t)\,\xi(t)$ where σ is an amplitude function modifying the white noise.

Hence,

$$\frac{dy}{dt} = \phi(y,t) + \sigma(y,t)\,\xi(t). \tag{2.7.3}$$

As we have seen before (equation (2.3.3)),

$$\sigma(y,t)\,\xi(t) = \sigma(y,t)\frac{dB}{dt} \tag{2.7.4}$$

where, $B(t)$ = the standard Wiener process.

Therefore,

$$\frac{dy}{dt} = \phi(y,t) + \sigma(y,t)\frac{dB}{dt}, \tag{2.7.5}$$

$$dy = \phi(y,t)d\phi + \sigma(y,t)dB. \tag{2.7.6}$$

In general, $\phi(y,t)$ and $\sigma(y,t)$, could be stochastic processes. This equation is called a stochastic differential equation (SDE) driven by Wiener process. Once the Wiener process enters into equation (2.7.4), y becomes a stochastic process, $Y(t,\omega)$, and in the differential notation SDE is written as

$$dY(t) = \phi(Y(t),t)dt + \sigma(Y(t),t)\,dB(t). \tag{2.7.7}$$

This actually means,

$$Y(t) = Y(0) + \int_o^t \phi(Y(t),t)\,dt + \int_o^t \sigma(Y(t),t)\,dB(t). \tag{2.7.8}$$

If we can find a function of Wiener process in the form of an Ito process that satisfies the above integral equation (2.7.8), we call that function a strong solution of SDE.

Strong solutions do not depend on individual realizations of Brownian motion. In other words, all possible realizations of an Ito process, which is a strong solution of a SDE, satisfy the SDE under consideration. Not all the SDEs have strong solutions. Other class of solutions is called weak solutions where solution to each individual realization is different from each other. In this section we will focus only on strong solutions. In many situations, finding analytical solutions to SDEs is impossible and therefore we will review a minimum number of SDEs and their solutions in order to facilitate the discussion in the subsequent chapters.

If $X(t)$ is a stochastic process and another stochastic process $Y(t)$ is related to $X(t)$ through the stochastic differential,

$$dY(t) = Y(t)\, dX(t) \quad , \tag{2.7.9}$$

with $Y(0) = 1$.

Thus $Y(t)$ is called the stochastic exponential of $X(t)$. If $X(t)$ is a stochastic process of finite variation thus the solution to equation (2.7.9) is,

$$Y(t) = e^{X(t)}, \tag{2.7.10}$$

and, for any process $X(t)$,

$Y(t) = e^{\xi(t)}$ satisfies the stochastic differential given above when

$$\xi(t) = X(t) - X(0) - \frac{1}{2}[X, X](t). \tag{2.7.11}$$

$[X, X](t)$ is quadratic variation of $X(t)$ and for a continuous function with finite variation $[X,X](t) = 0$.

For example, consider the following stochastic differential equation in differential form,

$$dX(t) = X(t)\, dB(t). \tag{2.7.12}$$

This SDE does not have a drift term and the diffusion term is an Ito integral.

We know, $[B, B](t) = t$.

Therefore from the above result,

$$\begin{aligned} \xi(t) &= B(t) - B(0) - \frac{1}{2}t, \\ &= B(t) - \frac{1}{2}(t). \end{aligned} \tag{2.7.13}$$

Then the solution to the SDE is

$$X(t) = e^{B(t) - \frac{1}{2}t}. \tag{2.7.14}$$

Now let us consider a similar SDE with a drift term:

$$dX(t) = \alpha\ X(t)\ dt + \beta\ X(t)\ dB(t)\ , \tag{2.7.15}$$

where α and β are constants.

Dividing it by $X(t)$,

$$\frac{dX(t)}{X(t)} = \alpha\ dt + \beta\ dB(t)\ . \tag{2.7.16}$$

This differential represents,

$$\begin{aligned} \int_0^t \frac{dX}{X(t)} &= \int_0^t \alpha\, dt + \int_0^t \beta\, dB(t), \\ &= \alpha\, t + \beta(B(t) - B(0)), \\ &= \alpha\, t + \beta\, B(t). \end{aligned} \tag{2.7.17}$$

The second term on the right hand side comes from Ito integration.

Now let us assume

$$\phi\,(t) = \alpha\, t + \beta\, B(t)\ . \tag{2.7.18}$$

Then the SDE becomes,

$$\int_0^t \frac{dX(t)}{X(t)} = \phi\,(t),$$

and

$$\xi(t) = \phi(t) - \phi(0) - \frac{1}{2}[\phi,\phi](t)\ .$$

$$\begin{aligned} [\phi,\phi](t) &= [(\alpha\, t + \beta\, B(t)),(\alpha\, t + \beta\, B(t))](t), \\ &= [\alpha\, t, \alpha\, t](t) + 2\alpha\ \beta\,[t, B(t)](t) + \beta^{\,2}\,[B,B](t), \\ &= 0 + 0 + \beta^2\ t. \end{aligned}$$

Therefore $\xi(t) = \alpha\, t + \beta B(t) - 0 - \frac{1}{2}\beta^{\,2}\, t$.

Then the solution to the SDE is

$$X(t) = \exp((\alpha - \frac{1}{2}\beta^{\,2})t + \beta\ B(t))\ .$$

Let us examine whether the stochastic process

$$X(t) = \exp((\alpha - \frac{1}{2}\beta^2)t + \beta\ B(t)). \tag{2.7.19}$$

is a strong solution to the differential equation

$$dX(t) = \alpha X(t)dt + \beta\, X(t)\, dB(t)\,.$$

We will define a function,

$$f(x,t) = \exp((\alpha - \frac{1}{2}\beta^2)t + \beta\, X).$$

$$X(t) = f(B(t),t),$$
$$= \exp((\alpha - \frac{1}{2}\beta^2)t + \beta\, B(t)).$$

We need to apply Ito formula for the two Ito processes $X_1(t)$ and $X_2(t)$.

$X_1(t) = B(t)$; $X_2(t) = t$ (a continuous function with finite variation);

$$dX_1.dX_2(t) = d[X_1, X_2] = 0\,;\ \ (dX_1)^2 = dt\,;\ \ \ (dX_2)^2 = 0\,.$$

$$\frac{\partial f}{\partial x} = \beta \exp((\alpha - \frac{1}{2}\beta^2)t + \beta x)\ \ ,$$

$$\frac{\partial^2 f}{\partial x^2} = \beta^2 \exp((\alpha - \frac{1}{2}\beta^2)t + \beta x)\ \ ,$$

$$\frac{\partial f}{\partial t} = (\alpha - \frac{1}{2}\beta^2)\exp((\alpha - \frac{1}{2}\beta^2)t + \beta x)$$
$$= (\alpha - \frac{1}{2}\beta^2)\exp((\alpha - \frac{1}{2}\beta^2)t + \beta x)$$

.

From Ito formula,

$$d(f(X_1, X_2))$$
$$= \frac{\partial f}{\partial x}dB(t) + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}dt + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(0) + \frac{1}{2}\frac{\partial^2 f}{\partial x\,\partial t}(0),$$
$$= \beta\exp((\alpha - \frac{1}{2}\beta^2)t + \beta\, B(t)) + (\alpha - \frac{1}{2}\beta^2)\exp(\alpha - \frac{1}{2}\beta^2)dt + \frac{1}{2}\beta^2\exp(\alpha - \frac{1}{2}\beta^2)dt.$$

$$d(X(t)) = d(f(B(t),t))$$
$$= \alpha\, \exp((\alpha - \frac{1}{2}\beta^2)t + \beta\, B(t))dt + \beta\, \exp((\alpha - \frac{1}{2}\beta^2)t + \beta\, B(t)dB(t),$$
$$= \alpha X(t)dt + \beta X(t)\, dB(t).$$

This proves that $X(t) = f(B(t),t)$ is a strong solution of the SDE given by equation (2.7.19).

We can see that if we can find a function $f(x,t)$, and for a given Wiener process $B(t)$, $X(t) = f(B(t),t)$ is a solution to the SDE of the form

$$dX(t) = \mu\,(X(t),t)dt + \sigma\,(X(t),t)\, dB(t)\,. \qquad (2.7.20)$$

$X(t)$ should also satisfy,

$$X(t) = X(0) + \int_0^t \mu(X(s), s) + \int_0^t \sigma \, dB(s) \, ,$$

provided that $\int_o^t \mu \; ds$ and $\int_o^t \sigma \, ds(s)$ exist.

Solution to the general linear SDE of the form,

$$dX(t) = (\alpha(t) + \beta(t)X(t))dt + (\gamma(t) + \delta(t)X(t)) \, dB(t). \tag{2.7.21}$$

where α, β, γ and δ are given adapted processes and continuous functions of t, can be quite useful in applications.

The solution can be expresses as a product of two Ito processes (Klebaner, 1998)

$$X(t) = u(t) \, v(t) \, , \text{ where ,}$$

$$du(t) = \beta \, u(t)dt + \delta u(t) \, dB(t) \, , \text{ and}$$

$$dv(t) = a \, dt + b \, dB(t) \, .$$

$u(t)$ can be solved by using a stochastic exponential as shown above and once we have a solution, we can obtain $a(t)$, $b(t)$ by solving the following two equations:

$$b(t) \, u(t) = \gamma(t) \, , \text{ and}$$

$$a(t) \, u(t) = \alpha(t) - \delta(t) \, \gamma(t) \, .$$

Then the solution to the general linear SDE is given by (Klebaner, 1998) :

$$X(t) = u(t)\left(X(0) + \int_0^t \frac{\alpha(s) - \delta(s) \, \gamma(s)}{u(s)} ds + \int_0^t \frac{\gamma(s)}{u(s)} dB(s) \right) \tag{2.7.22}$$

As an example let us solve the following linear SDE:

$$dx(t) = a \, X(t) \, dt + dB(t) \, , \tag{2.7.23}$$

where a is a constant.

Here $\beta(t) = a, \; \gamma(t) = 1, \; \alpha(t) = 0$, and $\delta(t) = 0$.

Using the general solution with

$$\begin{aligned} du(t) &= a \, u(t) \, dt + (0) \, dB(t), \\ &= a(t) \, u(t) \, dt. \end{aligned}$$

From stochastic exponential,

$$u(t) = \exp(a \, t) \quad .$$

Therefore,

$$X(t) = \exp(at)(X(o) + \int_o^t \exp(-as)\ \ dB(s)) .$$

This is also the solution of the SDE given by equation (2.7.23).

The integral in the solution given above is an Ito integral and should be calculated according Ito integration. For non-linear stochastic differential equations, appropriate substitutions may be found to reduce them to linear ones.

2.8 The Estimation of Parameters for Stochastic Differential Equations Using Neural Networks

Stochastic differential equations (SDEs) offer an attractive way of modelling the random system dynamics, but the estimation of the drift and diffusion coefficients remains a challenging problem in many situations. There are various statistical methods that are used to estimate the parameters in differential equations driven by Wiener processes. In this section we offer an alternative approach based on artificial neural networks to estimate the parameters in a SDE. Readers who are familiar with neural networks may skip this section. Artificial Neural Networks (ANNs) as discussed in chapter 1 are universal function approximators that can map any nonlinear function, and they have been used in a variety of fields, such as prediction, pattern recognition, classification and forecasting. ANNs are less sensitive to error term assumptions and they can tolerate noise and chaotic behaviour better than most other methods. Other advantages include greater fault tolerance, robustness and adaptability due to ANNs' large number of interconnected processing elements that can be trained to learn new patterns (Bishop, 1995). The Multilayer Perceptron (MLP) network is among the most common ANN architecture in use. It is one type of feed forward networks wherein the connections are only allowed from the nodes in layer i to the nodes in layer i+1. There are other more complex neural network architectures available, such as recurrent networks and stochastic networks; however MLP networks are always sufficient for dealing with most of the recognition and classification problems if enough hidden layers and hidden neurons are used (Samarasinghe, 2006). We show how to use the output values from the SDE solutions of the equations to train neural networks, and use the trained networks to estimate the SDE parameters for given output data. MLP networks will be used to solve this type of mapping problem.

The general form of SDE can be expressed by

$$dy(t) = \mu(y,t,\theta)\,dt + \sigma(y,t,\theta)\,dw(t) \tag{2.8.1}$$

where $y(t)$ = the state variable of interest, θ = a set of parameters (known and unknown), and $w(t)$ = a standard Wiener process. In practice, to determine the parameter θ, the system output variable y is usually observed at discrete time intervals, t, where $0 \le t \le T$, at M independent points: $y = \{y_1, y_2, ..., y_M\}$. Observed data are recorded in discrete time intervals, regardless whether the model is described best by a continuous or discrete intervals.

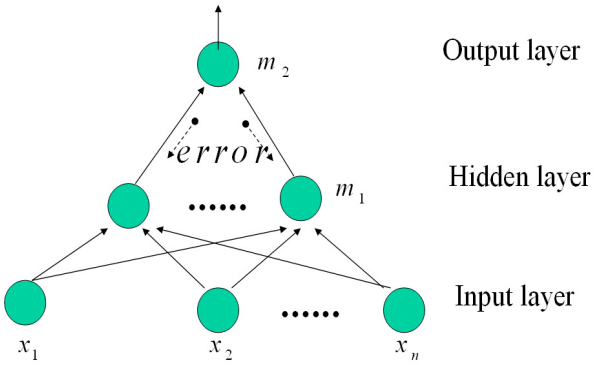

Figure 2.8. Basic structure of a MLP with backpropagation algorithm.

A MLP shown in Figure 2.8 has one hidden layer (m_1) and one output layer (m_2), and all the layers are fully connected to their subsequent layer. Connections are only allowed from the input layer to the hidden layer, and then, from the hidden layer to the output layer.

Rumelhart (1986) developed the backpropagation learning algorithm and it is commonly used to train ANNs due to its advanced ability to generalize wider variety of problems. A typical backpropagation learning algorithm is based on an architecture that contains a layer of input neurons, output neurons, and one or more hidden layers; these neurons are all interconnected with different weights. In the backpropagation training algorithm, the error information is passed backward from the output layer to the input layer (Figure 2.8). Weights are adjusted with the gradient descent method.

The ANN is trained by first setting up the network with all its units and connections, and then initialising with arbitrary weights. Then the network is trained by presenting examples. During the training phase the weights on connections change enabling the network to learn. When the network performs well on all training examples it should be validated on some other examples that it has not seen before. If the network can produce reasonable output values which are similar to validation targets and contain only small errors, it is then considered to be ready to use for problem solving.

Both linear and nonlinear SDEs are examined in this section. The linear SDE (Eq. (2.8.2)) is expressed by a one-dimensional diffusion equation. Its drift term has a linear relationship to the output variable of the model, and the diffusion term represents the noise in the model. Eq. (2.8.3) is arbitrarily chosen as a representative nonlinear SDE:

$$dX(t) = \alpha X(t)dt + \gamma X(t)dw(t), \text{ and} \tag{2.8.2}$$

$$dX(t) = \alpha X(t)dt + \beta X^2(t)dt + \gamma dw(t), \tag{2.8.3}$$

where a, β = constant coefficients to be estimated as parameters, and γ = a constant coefficient to adjust the noise level (amplitude).

For each particular parameter a or the combination of parameters a and β, we can generate one realisation of SDE output through Eq. (2.8.2) or (2.8.3). The range of α and β used in

the experiments is assumed to be [1, 2]. In addition, γ is used to adjust the proportion of the contribution of diffusion term, and the range of γ is [0.01, 0.1] in the linear case and [0.5, 3] in the nonlinear case.

The discrete observations $X(t)$ of these two equations are obtained at the sampling instants. Suppose the number of samples to be t_n, we consider the first t_n time steps starting from $X_0 = 1$ and the size of sampling interval Δt = 0.001. All the values come from the solution of SDEs. It has been shown that using Ito formula, Eq. (2.8.2) has an analytical solution (section 2.7),

$$X(t) = X_0 \exp\left[(\alpha - \frac{\gamma^2}{2})t + \gamma w(t) \right]. \tag{2.8.4}$$

For Eq. (2.8.4), we use the Euler method for the numerical solution. The numerical solution of $\gamma = 0$ has been compared with the analytical solution of the equation.

Before we describe the neural networks data sets, we clarify the terminology about "training", "validation" and "test" data sets. In the literature of machine learning and neural networks communities, different meaning of the words "validation" and "test" are employed. We restrict ourselves to the definitions given by Ripley (1996): a training set is used for learning, a validation set is used to tune the network parameters and a test set is used only to assess the performance of the network.

We generate a number of SDE realisations for a specified range of parameters with some patterns of Wiener processes to train the ANN. These data sets are called training data sets and validation sets are randomly chosen from the training sets. In order to test the prediction capability of the ANN, test data sets are generated with different patterns of Wiener processes within the same range of parameters as the training data sets.

Obviously if the test data sets were generated from SDEs which contain only a single Wiener process, the result would be biased if this Wiener process was coincidently similar to the one used to generate the training data sets. To fairly assess the performance of networks, five different patterns of Wiener processes are used to generate the test data sets.

To determine the value of time step t_n, we have taken different t_n values, where $t_n^{\min}$ = 10 to $t_n^{\max}$ = 200 and Δt_n = 10, to generate the training and test data sets. We found that 50 values were sufficient to represent the pattern of SDEs in order to train neural networks. Further increase in t_n did not increase the neural networks performance in parameter estimation. Therefore 50 time steps are used in our computational experiments.

All the experiments are carried out on a personal computer running the Microsoft Windows XP operating system. We use a commercial ANN software, namely NeuroShell2, for the neural network computations. It is recommended for academic users only, or those users who are concerned with classic neural network paradigms like backpropagation. Users interested in solving real problems should consider the NeuroShell Predictor, NeuroShell Classifier, or the NeuroShell Trader (Group, W.S., 2005).

Among all the parameters in MLP, the numbers of input and output neurons are the easiest parameters to be determined because each independent variable is represented by its own input neuron. The number of inputs is determined by the number of sampling instants in the SDE's solution, and the number of outputs is determined by the number of parameters which need to be predicted. In terms of the number of hidden layers and hidden layer neurons, we try a network that started with one layer and a few neurons, and then test different hidden layers and neurons to achieve the best ANN performance. In the following experiments, (X – Y – Z) is used to denote to the networks, where X is the number of input nodes, Y is the number of hidden nodes and Z is the number of output nodes.

We found that the logistic function was always superior to other five transfer functions used in NeuroShell2, logistic, linear, hyperbolic tangent function, Sine and Gaussian, as input, output and hidden layer functions because of its nonlinear and continuously differentiable properties, which are desirable for learning complex problems. In addition to the logistic function, we use the default values of 0.1 in NeuroShell2 for both learning rate and momentum as we found that it was always appropriate.

The number of training epochs plays an important role in determining the performance of the ANN. An epoch is the presentation of the entire training going through the network. ANNs need sufficient training epochs to learn complex input-output relationships. However excessive training epochs require unnecessarily long training time and cause over fitting problems where the network performs during the training very well but fails in testing (Caruana, 2001). To monitor the over fitting problem, we set up 20% of the training sets as validation sets and and the ANN monitors errors on the validation sets during training. The profile of the error plot for the training and validation sets during the training procedure indicates whether further training epoch is needed. We can stop training when the error of the training set plot keeps decreasing but that of the validation set plot has an increasing or flat line at the end.

In order to test the robustness of neural networks, we need to measure the level of noise in the diffusion term of a SDE with respect to its drift term. Thus the diffusion parameter γ is used to adjust the noise level. The higher γ value indicates greater noise and increases the influence of the contribution of the diffusion term to the entire solution. As one can assume, the increased noise levels raises the difficulty of estimation. To measure it, we define P_γ for linear equation (2.8.2) as

$$P_\gamma \equiv \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\gamma\, dw(i)}{\alpha\, dt}\right|,$$

and for nonlinear equation (2.8.3),

$$P_\gamma \equiv \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\gamma\, dw(i)}{\alpha\, dt\, x(i) + \beta\, dt\, x(i)^2}\right|, \tag{2.8.5}$$

where n = the number of time steps, and dt = time differential.

There are two parameters, α and β, in the drift term of the nonlinear SDE. We define P_a to determine the strength of the linear term (i.e. $\alpha\, dt\, x(t)$). Similarly, P_β indicates the measurement of strength of nonlinear term. They can be defined as

$$P_\alpha \equiv \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\alpha\, dt}{\alpha\, dt + \beta\, dt\, x(i)}\right|, \tag{2.8.6}$$

and

$$P_\beta \equiv \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\beta\, dt\, x(i)}{\alpha\, dt + \beta\, dt\, x(i)}\right|. \tag{2.8.7}$$

R^2 (coefficient of multiple determinations) is a statistical indication of data sets which is determined by multiple regression analysis, and it is an important indicator of the ANN performance used in NeuroShell2 (Triola, 2004). R^2 compares the results predicted by ANN with actual values, and it is defined by

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(y_i - \overline{y})^2}, \tag{2.8.8}$$

where y = actual value, $\hat{y}$ = predicted value of y, $\overline{y}$ = mean of y, and m = number of data patterns. In the case of parameter estimation for the linear SDE, one R^2 value is obtained for determining the accuracy of the predicted parameter a. For the nonlinear SDE, two R^2 values are calculated for determining the accuracy of the predicted parameters a and β. If the ANN predicts all the values correctly as the actual values, a perfect fit would result in an R^2 value of 1. In a very poor fit, the R^2 value would be close to 0. If ANN predictions are worse than the mean of samples, the R^2 value will be less than 0.

In addition to R^2, the Average Absolute Percentage Error (*AAPE*) is also used for evaluating the prediction performance where needed (Triola, 2004). The *AAPE* can be defined as,

$$AAPE = \frac{1}{m}\sum_{i=1}^{m}\frac{|y_i - \hat{y}_i|}{y_i}100 \tag{2.8.9}$$

where y = target value, $\hat{y}$ = predicted value of y, and m = number of data patterns.

The performance of ANN is evaluated by assessing the accuracy of the estimated parameters. Different ANN architectures including various combinations of hidden layers, neurons, and training epochs are used to obtain the optimum neural network. Further, a range of diffusion term is used to evaluate the effect of different level of stochasticity on the performance of ANN.

We do not have *a priori* knowledge of the optimal ANN architecture at first; therefore we choose the default parameters in NeuroShell2 for one hidden layer MLP network, which has

68 hidden neurons and the logistic transfer function for the hidden layers and output layer. In addition, α = [1, 2], $\Delta\alpha$ = 0.01 and γ = 0.03 are used for the parameters of SDE.

The experiments show that when training data set is developed using only one Wiener process, over fitting problem is obvious. The average error in the training set continues to decrease during training process. Because of the powerful mapping capability of neural networks, the average error between the target and network outputs approaches zero as the training continues. During the first four epochs, the average error in the test set drops significantly. It reaches the lowest at the epoch 8. After that, the validation set error starts rising although the training set error is getting smaller. The reason for this increasingly poor generalization is that the neural network tends to track every individual point in the training set created by a particular pattern of Wiener process, rather than seeing the whole character of the equation.

When the training data set is produced by more than one Wiener process, over fitting significantly decreases. The average error in the validation set continues to drop and remains stable after certain epochs. We examine the relationship between the number of Wiener processes and ANN prediction ability.

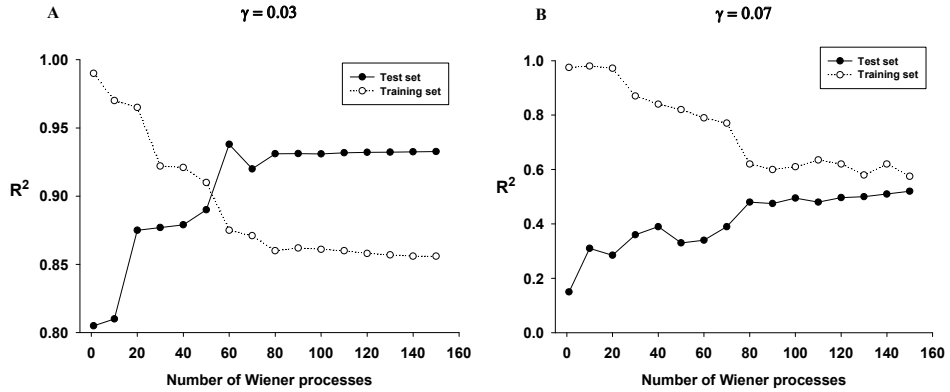

Figure 2.9. R^2 on the training and test sets against the number of Wiener Processes used to produce the training sets in the case of γ = 0.03 (A) and γ = 0.07 (B).

The same ANN architecture and SDE parameters as the previous section are used here. Additionally we test a set of noisier data with γ = 0.07. The results are obtained with the same numbers of training epochs. Figure 2.9 shows the influence of the number of Wiener processes that are used to produce training data sets. It indicates that as the number of Wiener Processes in the training sets increases, the network produces higher R^2 values for the test sets. It should be noted that the size of training data set expands as more Wiener processes are employed, and consequently the expansion causes slower training. Therefore, although there is a marginal improvement on R^2 value when more than 80 Wiener processes are used, we limit the number of Wiener Processes to 100 in further investigations.

We use the same SDE parameters except γ = 0.07 to create training and test data sets, and 100 Wiener processes are used to produce the training data sets. All the R^2 values are obtained by using early stopping. The results in Table 2.1 suggest that when there is only one hidden layer and the number of neurons in the hidden layer is very small, the performance of the network is poor because the network does not have enough "power" to learn the input-output relationship. When the number of neurons in the hidden layer is close to the half number of input neurons, the performance reaches a higher accuracy. Further increase in the number of hidden layers and neurons does not improve the performance.

The ANN performance is investigated for different combinations of drift and diffusion terms. We use three different MLP architectures, 50-30-1, 50-15-15-1, and 50-10-10-10-1, to train and test the data sets, and record the best performance.

The ANN performance is investigated for different combinations of drift and diffusion terms. We use three different MLP architectures, 50-30-1, 50-15-15-1, and 50-10-10-10-1, to train and test the data sets, and record the best performance.

1 hidden layer	Test set R^2	2 hidden layers	Test set R^2	3 hidden layer	Test set R^2
50-3-1	0.2728	50-3-3-1	0.5103	50-3-3-3-1	0.4920
50-10-1	0.5222	50-5-5-1	0.4916	50-5-5-5-1	0.4576
50-30-1	0.5392	50-15-15-1	0.5125	50-10-10-10-1	0.5075
50-50-1	0.5151	50-25-25-1	0.4986	50-20-20-20-1	0.4987
50-100-1	0.4980	50-50-50-1	0.4936	50-30-30-30-1	0.5072
50-200-1	0.4969	50-100-100-1	0.4669	50-100-100-100-1	0.4892

Table 2.1. R^2 variation on test set with different hidden neurons and hidden layers.

Figure 2.10A demonstrates that the ANN performance decreases as the magnitude of the diffusion term increases and Figure 2.10B shows that the target and network output in the test sets are in good agreement when γ = 0.01. Because the test set is created by 5 Wiener processes, it should be noticed that there are five repetitive sub-data sets and each of them represents a range of α values, which is from 1 to 2, with one pattern of Wiener process. By observing the sub-data sets separately, we can gain a better understanding on how noise influences the estimation of the parameter. As the γ value reaches 0.05 (Figure 2.10C) and the ratio of diffusion term and drift term reaches 0.67 (shown in Figure 2.10A), the 2nd, 3rd and 5th sub-data sets show more accurate predictions than the 1st and 4th sets. Figure 2.10D demonstrates that the network-generated outputs just tend to use the average of targets in most of the sub-data sets when γ = 0.10 where the weight of diffusion term is more than that of the drift term (P_γ = 1.39 as shown in Figure 2.10A).

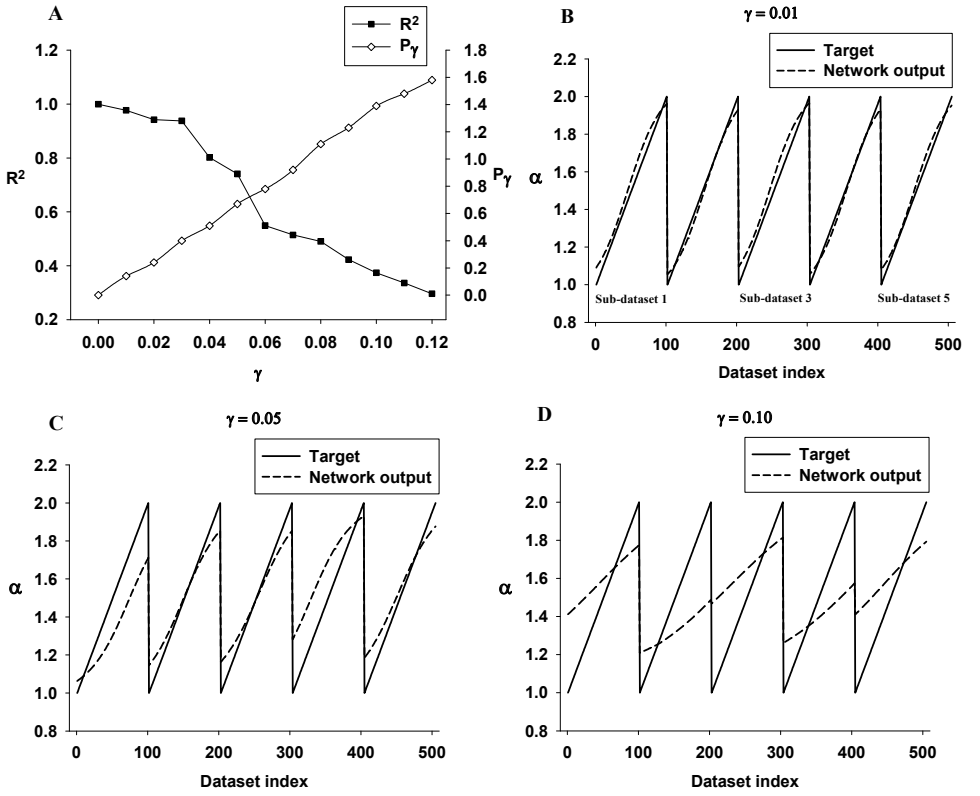

Figure 2.10. A: The neural network performance decreases as the diffusion term in SDEs increases. B, C and D: Target values and network outputs when γ= 0.01, 0.05 and 0.10; x-axis represents the index in testing datasets where five Wiener processes were used.

We investigate nonlinear SDE as well. Moreover, because the nonlinear SDE contains two parameters, we investigate how the accuracy of estimation varies for different combination of parameters. The parameter values and ranges of the SDE are as follows: α = [1, 2], $\Delta\alpha$ = 0.05, β = [1, 2], $\Delta\beta$ = 0.05 and γ= 0.5. We use early stopping to find out the best results. From Table 2.2, the different network architectures result in a very similar performance. The R^2 values for α are very close to zero while the R^2 values for β are all more than 0.9. According to the statistical meaning of R^2 given previously, we consider that the neural networks fail to predict a and succeed in predicting β. We explore the reason for the difference between a and β later.

Network architecture	$R^2(\alpha)$	$R^2(\beta)$	Network architecture	$R^2(\alpha)$	$R^2(\beta)$
50-10-2	0.0256	0.9161	50-10-10-2	-0.0135	0.9209
50-30-2	0.0349	0.9453	50-20-20-2	-0.0183	0.9299
50-60-2	0.0395	0.9406	50-50-50-2	0.0134	0.9310
50-100-2	0.0296	0.9209	50-10-10-10-2	0.0128	0.9198
50-200-2	0.0354	0.9299	50-50-50-50-2	-0.0164	0.9257

Table 2.2. Network performance in the nonlinear SDE as network architecture changes.

We use three network architectures, 50-30-2, 50-60-2 and 50-50-50-2, to estimate parameters for different SDEs and recorded the best results. The results in Table 2.3 indicate that the accuracy of network performance decreases as the strength of diffusion terms in SDEs increases, which is similar to the linear equation. Figure 2.11 shows that comparing with the results in the linear case (Figure 2.10A), the prediction capability of networks for the nonlinear case is poorer due to the complexity of input-output relationship in the nonlinear SDEs.

Range of a	Range of β	γ	P_α	P_β	P_γ	$R^2(a)$	AAPE (α)	$R^2(\beta)$	AAPE (β)
[1,2]	[1,2]	0.5	0.12	0.88	0.10	0.0349	17.79	0.9453	4.02
[1,2]	[1,2]	2	0.11	0.89	0.40	0.0006	18	0.6833	9.92
[1,2]	[1,2]	3	0.11	0.89	0.60	-0.012	17.59	0.3687	12.96

Table 2.3. Network performance in the nonlinear SDE as diffusion term increases.

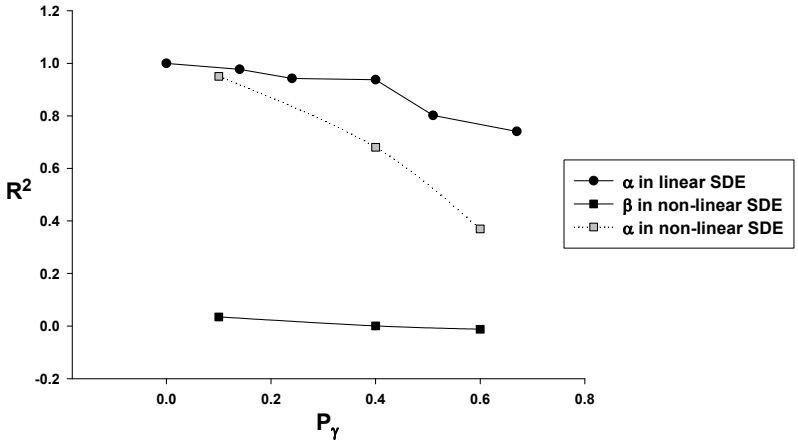

Figure 2.11. The scattered graph of R^2 values for the parameters in the linear and the nonlinear equations against their corresponding P_γ values.

Range of a	Range of β	γ	P_α	P_β	P_γ	$R^2(\alpha)$	AAPE (α)	$R^2(\beta)$	AAPE (β)
[1,2]	[0.2,0.4]	0.35	0.47	0.53	0.39	0.2932	13.91	0.3891	12.98
[1,2]	[1,2]	2	0.11	0.89	0.40	0.0006	18	0.6833	9.92
[0.8,1.6]	[0.02,0.04]	0.15	0.89	0.11	0.42	0.7735	8.46	0.0108	17.94

Table 2.4. Network performance for different parameters in the nonlinear SDE

To investigate the reason for largest differences in R^2 values for a and β, we change the magnitudes of α term and β term in SDEs by altering parameters α and β values while keeping diffusion level an approximate constant. Table 2.4 shows that the bigger the contribution of a term containing a particular parameter (P_a or P_β), the smaller the error (*AAPE*) and better the prediction (R^2) for that parameter. Therefore, we conclude that the accuracy of a parameter in a nonlinear SDE is dependent on its term that contributes *pro rata* to the drift term.

In the data preparation stage, we use different time steps to solve SDEs and found 50 data points are sufficient to represent the realisation of SDEs. In addition, we emphasise the effect of the number of Wiener processes used to create training data sets. Increasing the number of Wiener processes boosts the performance of networks considerably and eliminates the over fitting problem. When over fitting occurs, the resulting network is accurate on the training set but perform poorly on the test set. When the number of Wiener processes used to generate training data sets is increased, the learning procedure finds common features amongst the training sets that enable the network to correctly estimate the parameter(s) in test data sets.

In the ANN training procedure, we use early stopping to obtain the optimum test results. We also employ different MLP architectures, transfer functions, learning rates and momentums. However we find that these factors do not increase the performance of ANNs significantly.

The diffusion level in a SDE has a significant impact on the network performance. In the linear SDE, when the ratio of diffusion term and drift term is below 0.40, the network can estimate the parameter accurately (R^2>0.93). When the ratio reaches 0.67, the network estimates the parameter accurately only when Wiener processes in test sets and in training sets are similar. If the diffusion term is larger than the drift term, the network cannot predict the parameter(s) and only tends to give an average value of the parameters used for training datasets. For nonlinear SDEs, the estimation ability of a network is generally poorer than that for the linear SDEs. Furthermore, the accuracy of a parameter in a nonlinear SDE is dependent on its term that contributes *pro rata* to the drift term.

We can conclude that the classical neural networks method (MLP with backpropagation algorithm) provides a simple but robust parameter estimation approach for the SDEs that are under certain noisy conditions, but this estimation capability is limited for the SDEs having a high diffusion level. When the diffusion level is high (>10%-20%), the statistical methods also fail to estimate parameters accurately.