Numerical Methods to Solve Stochastic Differential Equations

Anjali Maria Dhanpal

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

April 2021

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering.

Abstract

The thesis presents an introduction to numerical methods to solve Stochastic Differential Equations by comparison of methods from an application point of view. These numerical methods are used in the simulation of a power system based model from a paper by Professor Xiaozhe Wang and Konstantin Turitsyn [1]. The methods are compared by simulating the trajectories of the discretization and the diffusion.

Résumé

La thèse présente une introduction aux méthodes numériques pour résoudre les équations différentielles stochastiques par comparaison de méthodes d'un point de vue applicatif. Ces méthodes numériques sont utilisées dans la simulation d'un modèle basé sur un système électrique à partir d'un article du professeur Xiaozhe Wang et Konstantin Turitsyn cite wang. Les méthodes sont comparées en simulant les trajectoires de la discrétisation et de la diffusion.

Acknowledgment

I would like to express my gratitude to several individuals who constantly helped me throughout my academic life at McGill.

Firstly I thank Professor Hannah Michalska, my supervisor and a wonderful person. She guided and moulded me patiently through the entire research and helped shape this research thesis elegantly.

I am grateful to my friend and fellow research partner Rohitha Ratnam, who has been a great encouragement and advisor to solve many of the hurdles in the research as well as in this academic journey for the past two years.

I thank my family and friends who have been a constant support throughout my life.

Preface

This is to declare that the work presented in this document was completed and carried out by Anjali Maria Dhanpal guided by Professor Hannah Michalska. It presents the comparison of various numerical methods to solve Stochastic Differential Equations and it's application on a paper named PMU- Based Estimation of Jacobian Matrix by Professor Xioazhe Wang and Konstantin Turitsyn [1]. The theoretical background for Stochastic Differential Equations is based on the research notes by Särkkä [2], which is duly acknowledged.

Contents

1 Ordinary Differential Equations [2] 3
1.1 Solutions of General Linear Differential Equations 4
1.2 Numerical Solutions of Non-Linear Equations 5
1.3 Picard-Lindelöf theorem 6

2 Introduction to Stochastic Processes [4] 7
2.1 Motivation 7
2.2 Probability Spaces 9
2.3 Random Variables 10
2.4 Stochastic Process 11
2.5 Expected Value and Variance 11
2.6 Distribution Function 12

3 Stochastic Differential Equations [2] 13
3.1 Brownian Motion 13
3.2 Linear Differential Equations with Driving White Noise 16
3.3 Heuristic Solutions of Linear SDEs 17
3.4 Heuristic Solutions of Non-Linear SDEs 19
3.5 Existence and Uniqueness of Solutions 19

4 Statistics of SDEs [2] 21
4.1 Fokker-Planck-Kolmogorov Equation 21
4.2 Mean and Covariance of SDE 23
4.3 Higher Order Moments of SDEs 25
4.4 Mean and covariance of linear SDEs 25

5 Numerical Methods for Solving SDEs 28
5.1 Itô-Taylor series of SDEs [2] 28
5.2 Euler-Maruyama Method [2] 31

5.2.1 Properties of the Algorithm [2][8] 32
5.3 Milstein's Method [2] 33
5.3.1 Properties of the Algorithm [2] 35
5.3.2 Example 35
5.4 Itô-Taylor Method [2] 36
5.4.1 Properties of the Algorithm [2] 39

6 Application of the Numerical Methods for Solving SDEs **40**
6.1 Estimation of the Dynamic State Jacobian Matrix [1] 40
6.2 System Model [1] 42
6.3 Derivatives of Itô- Taylor Method for the System model 43
6.4 Euler-Maruyama Method 51
6.5 Milstein's Method 58
6.6 Ito-Taylor Method 64
6.7 Conclusion 71

A Itô-Calculus and SDEs [2] **74**
A.1 Taylor Series of ODEs 74
A.2 Itô Formula 79
A.3 Explicit Solutions of Linear SDEs 81

Chapter 1

Ordinary Differential Equations [2]

An ordinary differential equation (ODE) is an equation, where the unknown quantity is a function, and the equation involves derivatives of the unknown function. For example, the second order differential equation for a forced spring can generally be expressed as

$$\frac{d^2x(t)}{dt^2} + \gamma \frac{dx(t)}{dt} + \nu^2 x(t) = \omega(t) \tag{1.1}$$

where ν and γ are constants that determine the resonant angular velocity and damping of the spring. The force $\omega(t)$ is some given function which may or may not depend on time. In the Equation (1.1) the position variable x is called the dependent variable and time t is the independed variable. The equation is second order, because it contains the second derivative and it is linear, because $x(t)$ appears linearly in the equation. The equation is non-homogenous, because it contains the forcing term $\omega(t)$.

To solve the differential equation it is necessary to know the initial conditions. It means that we need to know the spring position $x(t_0)$ and velocity $dx(t_0)/dt$ at some fixed initial time t_0. Fixing some other boundary conditions instead of the initial conditions yields a unqiue solution to the differential equation.

Differential equations of order n can be converted into n first order vector differential equation. Considering the spring model Equation (1.1) and defining the state variables as $x(t) = (x_1, x_2) = (x(t), dx(t)/dt)$, Equation (1.1) can be written as

$$\underbrace{\begin{pmatrix} dx_1(t)/dt \\ dx_2(t)/dt \end{pmatrix}}_{dx(t)/dt} = \underbrace{\begin{pmatrix} 0 & 1 \\ -\nu^2 & -\gamma \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}}_{f(x(t))} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{L} \omega(t). \tag{1.2}$$

Equation (1.2) can be represented as

$$\frac{dx(t)}{dt} = f(x(t), t) + L(x(t), t)w(t), \tag{1.3}$$

where the vector valued function $x(t) \in \mathbb{R}^n$ is called the state of the system and $\omega(t) \in \mathbb{R}^s$ is a forcing function.

The spring model in Equation (1.2) can also be expressed as

$$\frac{dx(t)}{dt} = u(t)x(t) + v(t)\omega(t), \tag{1.4}$$

An nth order differential equation can always be converted into equivalent vector valued first order differential equations, called the state-space form.

1.1 Solutions of General Linear Differential Equations

Consider a general time-varying linear differential equation,

$$\frac{dx}{dt} = u(t)x, \text{ for a given } x(t_0), \tag{1.5}$$

The solution to Equation (1.5) can be implicitly expressed as

$$x(t) = \psi(t, t_0)x(t_0), \tag{1.6}$$

where $\psi(t, t_0)$ is the transition matrix with the following properties:

$$\begin{aligned} \partial\psi(\tau, t)/\partial\tau &= u(\tau)\psi(\tau, t) \\ \partial\psi(\tau, t)/\partial t &= -\psi(\tau, t)u(t) \\ \psi(\tau, t) &= \psi(\tau, s)\psi(s, t) \\ \psi(t, \tau) &= \psi^{-1}(\tau, t) \\ \psi(t, t) &= I \end{aligned} \tag{1.7}$$

Consider a non-homogeneous equation,

$$\frac{dx(t)}{dt} = u(t)x(t) + v(t)\omega(t), \text{ for a given } x(t_0), \tag{1.8}$$

Given the transition matrix, the solution to Equation (1.8) is

$$x(t) = \psi(t, t_0)x(t_0) + \int_{t_0}^{t} \psi(t, \tau)v(\tau)w(\tau)d\tau \tag{1.9}$$

Thus, linear equations can be solved analytically.

1.2 Numerical Solutions of Non-Linear Equations

For a general non-linear differential equation of the form

$$\frac{dx(t)}{dt} = f(x(t), t), \text{ for a given } x(t_0) \tag{1.10}$$

there is no general way to find an analytic solution. However, it is possible to approximate the solution numerically. Integrating the Equation (1.10) from t to $t + \Delta t$ yields,

$$x(t + \Delta t) = x(t) + \int_{t}^{t+\Delta t} f(x(\tau), \tau)d\tau \tag{1.11}$$

If the integral on the right hand side can be computed, the solution at time steps $t_0, t_1 = t_0 + \Delta t, t_2 = t_0 + 2\Delta t$ can be generated by iterating the Equation (1.11):

$$\begin{aligned} x(t_0 + \Delta t) &= x(t_0) + \int_{t_0}^{t_0+\Delta t} f(x(\tau), \tau)d\tau \\ x(t_0 + 2\Delta t) &= x(t_0 + \Delta t) + \int_{t0+\Delta t}^{t+2\Delta t} f(x(\tau), \tau)d\tau \\ x(t_0 + 3\Delta t) &= x(t_0 + 2\Delta t) + \int_{t0+2\Delta t}^{t+3\Delta t} f(x(\tau), \tau)d\tau \\ &\vdots \end{aligned} \tag{1.12}$$

It is now possible to derive various numerical methods by constructing approximations to the integrals on the right hand side. There exists a wide class of other numerical ODE solvers with fixed or variable step size, which can be found in [2] [3].

1.3 Picard-Lindelöf theorem

One important issue in differential equations is the question if the solution exists and whether it is unique. To analyze this question, consider a generic equation of the form

$$\frac{dx(t)}{dt} = f(x(t), t),\ x(t_0) = x_0, \tag{1.13}$$

where $f(x,t)$ is some given function. If the function $t \mapsto f(x(t), t)$ happens to be Riemann integrable, then we can integrate both sides from t_0 to t to yield

$$x(t) = x_0 + \int_{t_0}^{t} f(x(\tau), \tau) d\tau \tag{1.14}$$

Using this identity an approximate solution to the differential equation can be found by following Picard's iteration.

Picard's Iteration.
1. Start with initial guess $\varphi_0(t) = s_0$.
2. Compute approximations $\varphi_1(t), \varphi_2(t), \varphi_3(t), \cdots$ via the following recursion

$$\varphi_{n+1}(t) = x_0 + \int_{t_0}^{t} f(\varphi_n(\tau), \tau) d\tau \tag{1.15}$$

The iteration can be shown to converge to the unique solution

$$\lim_{n\to\infty} \varphi_n(t) = x(t) \tag{1.16}$$

provided that $f(x,t)$ is continuous in both arguments and Lipschitz continuous in the first argument.
The implication of the above is the Picard-Lindelöf theorem, which says that under the above continuity conditions the differential equation has a solution and it is unique at a certain interval around $t = t_0$. Emphasis on the function $f(x,t)$ which needs to be continuous. This is important because in the case of stochastic differential equations the corresponding function will be discontinuous everywhere and thus, a new existence theory is required for them.

In the following chapter, probability and stochastic processes are discussed first, followed by Stochastic Ordinary Differential Equations

Chapter 2

Introduction to Stochastic Processes [4]

2.1 Motivation

Fix a point $x_0 \in \mathbb{R}^n$ and consider the ordinary differential equation,

$$\frac{dx}{dt} = u(x(t)), \text{ for a given } x(0), \tag{2.1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a given smooth vector field, the solution is the trajectory $x(.) : [0, \infty) \longrightarrow \mathbb{R}^n$.

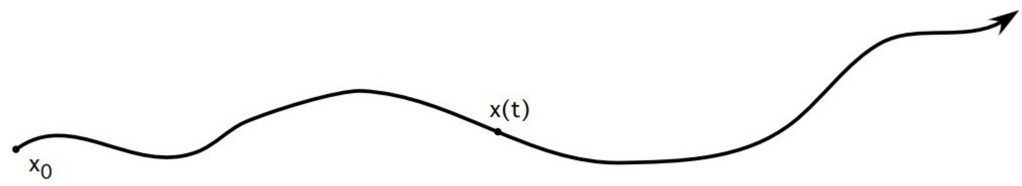

Figure 2.1 Trajectory of the Differential Equation

$x(t)$ is called the state of the system at time $t \geq 0$,

$$\dot{x}(t) := \frac{d}{dt}x(t) \tag{2.2}$$

In many applications, however, the experimentally measured trajectories of systems modeled by equation (2.2) do not behave as predicted.

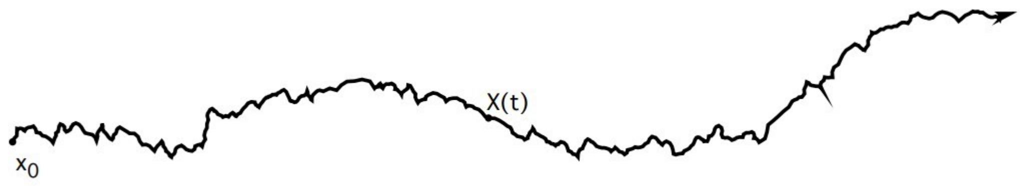

Figure 2.2 Sample path of the Stochastic Differential Equation

Thus, to accommodate the random effects disturbing the system, the equation (2.1) can be modified as,

$$\dot{x}(t) = f(x(t), t) + L(x(t), t)\xi(t), \forall t > 0 \text{ with } x(0) = x_0 \tag{2.3}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ and $L : \mathbb{R}^n \longrightarrow \mathbb{M}^{n\times m}$, $\mathbb{M}$ is the space of $n \times m$ matrices and $\xi(t)$ is called the "white noise".

Assume $m = n$, $f(x(t), t) = 0$ and $L(x(t), t) = I$, then the solution to equation (2.3) is the n-dimensional Wiener process or Brownian motion (refer Section (3.1), given by

$$\dot{W}(.) = \xi(.)$$

Thus, "white noise" is the time derivative of the Wiener process. The equation (2.3) can be written as,

$$dx(t) = f(x(t), t)dt + L(x(t), t)dW(t),\ x(0) = x_0 \tag{2.4}$$

The Equation (2.4) is interpreted as a Stochastic Differential Equation(SDE) and the solution is

$$x(t) = x_0 + \int_0^t f(x(\tau), \tau)d\tau + \int_0^t L(x(\tau), \tau)dW,\ \forall t > 0 \tag{2.5}$$

For $n = 1$, consider an SDE of the form

$$dx = f(x)dt + dW \tag{2.6}$$

with a solution $x(.)$. Suppose, $z : \mathbb{R} \longrightarrow \mathbb{R}$ is a smooth function defined as

$$Y(t) := z(x(t)),\ \forall t \geq 0 \tag{2.7}$$

From equation (2.6), solution to equation (2.7) is expected to be

$$dY = z'dx = z'f dt + z'dW \tag{2.8}$$

according to the usual chain rule, where $' = \frac{d}{dx}$.

$$dW \approx (dt)^2 \tag{2.9}$$

But the solution of equation (2.7) is computed as follows,

$$dY = z'dx + \frac{1}{2}z''(dx)^2 + ... \tag{2.10}$$

Substituting equation (2.6) and (2.9) into equation (2.10) yields,

$$\begin{aligned} dY &= z'(f dt + dW) + \frac{1}{2}z''(f dt + dW)^2 + ... \\ &= \left(z'f + \frac{1}{2}z''\right) dt + z'dW + \text{ higher order terms} \end{aligned} \tag{2.11}$$

The higher order terms can be neglected. Thus, the solution to equation (2.7) is given by

$$dY = \left(z'f + \frac{1}{2}z''\right)dt + z'dW \tag{2.12}$$

where "$\frac{1}{2}z''dt$" is an additional term that is not found in equation (2.8). Hence, ordinary calculus does not solve SDEs.

2.2 Probability Spaces

To understand stochastic processes, knowledge of probability and it's properties is necessary.

Terminology:

(i). A set $A \in \mathcal{U}$ is called an event; points $\omega \in \Omega$ are sample points.

(ii). $P(A)$ is the probability of the event A.

(iii). A property which is true except for an event of probability zero is said to hold almost surely (usually abbreviated "a.s.").

Consider a set denoted by Ω. Certain subsets of Ω are interpreted as "events".

Definition 2.2.1. A $\sigma - algebra$ is a collection $\mathcal{U}$ of subsets of Ω with the following properties:

(i). ϕ,$\Omega \in \mathcal{U}$.

(ii). If $A \in \mathcal{U}$, then $A^c \in \mathcal{U}$.

(iii). If A_1, A_2, ...$\in \mathcal{U}$, then

$$\bigcup_{k=1}^{\infty} A_k, \bigcap_{k=1}^{\infty} A_k \in \mathcal{U}$$

Here, $A^c := \Omega - A$ is the complement of A.

Definition 2.2.2. Let $\mathcal{U}$ be a $\sigma - algebra$ of subsets of Ω. $P : \mathcal{U} \longrightarrow [0, 1]$ is called a probability measure, provided:

(i). $P(\phi) = 0$, $P(\Omega) = 1$,

(ii). If A_1, A_2, ...$\in \mathcal{U}$, then

$$P(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} P(A_k)$$

(iii). If A_1,A_2, ... are disjoint sets in $\mathcal{U}$, then

$$P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k).$$

It follows that if A, $B \in \mathcal{U}$, then

$$A \subseteq B \text{ implies } P(A) \leq P(B).$$

Definition 2.2.3. A triple $(\Omega, \mathcal{U}, P)$ is called a probability space provided Ω is any set, $\mathcal{U}$ is a $\sigma - algebra$ of subsets of Ω, and P is a probability measure on $\mathcal{U}$.

2.3 Random Variables

The smallest σ-algebra containing all the open subsets of $\mathbb{R}^n$ is called Borel σ-algebra. $\mathcal{B}$ denotes the collection of Borel subsets of $\mathbb{R}^n$.

Definition 2.3.1. Let $(\Omega, \mathcal{U}, P)$ be a probability space. A mapping

$$X : \Omega \longrightarrow \mathbb{R}^n$$

is called an n - dimensional random variable, if for each $B \in \mathcal{B}$,

$$X^{-1}(B) \in \mathcal{U}.$$

It is equivalent to say that X is $\mathcal{U}$-measurable.

2.4 Stochastic Process

Definition 2.4.1. (i). A collection $X(t)|t \geq 0$ of random variables is called a stochastic process.

(ii). For each point $\omega \in \Omega$, the mapping $t \longrightarrow X(t, \omega)$ is the corresponding sample path.

If an experiment is conducted and the random values of $X(.)$ are observed as time evolves, it is the sample path $X(t, \omega)|t \geq 0$ for some fixed $\omega \in \Omega$. If the experiment is rerun, a different sample path is observed.

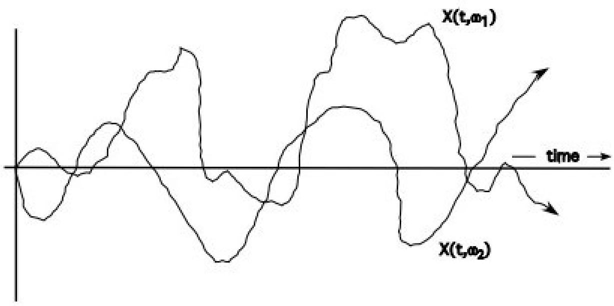

Figure 2.3 Two sample paths of a Stochastic Process

2.5 Expected Value and Variance

Definition 2.5.1. If $(\Omega, \mathcal{U}, P)$ is a probability space and X is a real-valued simple random variable, then

$$E(X) = \int_{\Omega} X dP$$

is the expected value (or mean value) of X.

Definition 2.5.2. If $(\Omega, \mathcal{U}, P)$ is a probability space and X is a real-valued simple random variable, then

$$V(X) := \int_\Omega |X - E(X)|^2 dP$$

is the variance of X.

It is observed that,

$$V(X) = E(|X - E(X)|^2) = E(|X|^2 - |E(X)|^2)$$

2.6 Distribution Function

Let $(\Omega, \mathcal{U}, P)$ be a probability space and $X : \Omega \longrightarrow \mathbb{R}^n$ be a random variable.
Notation: Let $x = (x_1, ..., x_n) \in \mathbb{R}^n$, $y = (y_1, ..., y_n) \in \mathbb{R}^n$. Then

$$x \leq y \text{ implies } x_i \leq y_i \ \forall i = 1, ..., n$$

Definition 2.6.1. (i). The distribution function of X is the function $F_X : \mathbb{R}^n \longrightarrow [0, 1]$ defined by

$$F_X(x) := P(X \leq x) \text{ for all } x \in \mathbb{R}^n$$

(ii). If $X_1, ..., X_m : \Omega \longrightarrow \mathbb{R}^n$ are random variables, their joint distribution function $F_{X_1,.....,X_m} : (\mathbb{R}^n)^m \longrightarrow [0, 1]$ is,

$$F_{X_1,.....,X_m}(x_1, ..., x_m) := P(X_1 \leq x_1, ..., X_m \leq x_m)$$

for all $x_i \in \mathbb{R}^n$, $i = 1, ..., m$.

Definition 2.6.2. Suppose $X : \Omega \longrightarrow \mathbb{R}^n$ is a random variable and $F = F_X$ is it's distribution function. If there exists a non-negative, integrable function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that

$$F(x) = F(x_1, ..., x_n) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_n} f(y_1, ..., y_n) dy_n ....dy_1$$

then f is called the density function for X.

It follows that

$$P(X \in B) = \int_B f(x)dx \text{ for all } B \in \mathcal{B}$$

The properties discussed in this chapter are helpful in understanding how an stochastic system functions. The next chapter discusses what an SDE is.

Chapter 3

Stochastic Differential Equations [2]

Stochastic Ordinary Differential Equation gives the realistic mathematical model of a systems/situation. The origin of the stochastic differential equations (SDEs) is dated back to the classic paper of Einstein [5], where a mathematical connection between microscopic random motion of particles and macroscopic diffusion equation is presented.

3.1 Brownian Motion

Brownian motion is critical in modelling a stochastic process as it represents 'white noise'. This motion enables modelling of systems which usually cannot be deterministically modeled, which means it has a property that is independent at each increment. The direction and magnitude at every change of the process is random and independent of the previous change.
Consider the two cases :
Case 1: Microscopic motion of Brownian particles
Let τ be a small interval and n is the number of particles suspended in liquid. During the time interval τ the x-coordinates of particles experience a change by displacement Δ. The number of particles within the displacements Δ and $\Delta + d\Delta$ is

$$dn = n\phi(\Delta)d\Delta, \tag{3.1}$$

where $\phi(\Delta)$ is the probability density of Δ. $\phi(n)$ can be assumed to be symmetric $\phi(\Delta) = \phi(-\Delta)$ and it differs from zero only for very small values of Δ.

Let $u(s,t)$ be the number of particles per unit volume. Then, the number of particles at time $t+\tau$ located at $x+dx$ is

$$u(x,t+\tau)dx = \int_{-\infty}^{\infty} u(x+\Delta,t)\phi(\Delta)d\Delta dx. \tag{3.2}$$

Because τ is very small, one can replace

$$u(x,t+\tau) = u(x,t) + \tau\frac{\partial u(x,t)}{\partial t}. \tag{3.3}$$

Expanding $u(x+\Delta,t)$ using Taylor series Δ :

$$u(x+\Delta,t) = u(x,t) + \Delta\frac{\partial u(x,t)}{\partial x} + \frac{\Delta^2}{2}\frac{\partial^2 u(x,t)}{\partial x^2} + \dots \tag{3.4}$$

Substituting into (3.3) and (3.4) into (3.2) gives

$$\begin{aligned} u(x,t) + \tau\frac{\partial u(x,t)}{\partial t} &= u(x,t)\int_{-\infty}^{\infty}\phi(\Delta)d\Delta + \frac{\partial u(x,t)}{\partial x}\int_{-\infty}^{\infty}\Delta\phi(\Delta)d\Delta \\ &\quad + \frac{\partial^2 u(x,t)}{\partial x^2}\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta + \dots \end{aligned} \tag{3.5}$$

where all the odd order terms vanish, recall that, $\int_{-\infty}^{\infty}\phi(\Delta)d\Delta = 1$.
Finally the obtained diffusion equation,

$$\frac{\partial u(x,t)}{\partial t} = D\frac{\partial^2 u(x,t)}{\partial x^2} \tag{3.6}$$

where,

$$\int_{-\infty}^{\infty}\frac{\Delta^2}{2}\phi(\Delta)d\Delta = D \tag{3.7}$$

This connection was significant during the time, because diffusion equation was only known as a macroscopic equation. Einstein was also able to derive a formula for D in terms of microscopic quantities. From this, the prediction for mean squared displacement of the particles as function of time is computed as:

$$z(t) = \frac{RT}{N}\frac{1}{3\pi\eta r}t, \tag{3.8}$$

where η is the viscosity of liquid, r is the diameter of the particles, T is the temperature, R is the gas constant, and N is the Avogadro constant.

Case 2: Langevin's model of Brownian motion

After Einstein's contribution to the Brownian motion, Langevin presented an alternative construction of Brownian motion leading to similar macroscopic properties, but more mechanical than Einstein's derivation.

Consider a small particle suspended in liquid, assuming two forces acting on the particle.
1. Friction force F_f, which by the Stokes law has the form:

$$F_f = -6\pi\eta r\nu, \tag{3.9}$$

where η is the viscosity, r is the diameter of particle and ν is the velocity.
2. Random force F_r caused by the random collision of the particles. From Newton's law,

$$m\frac{d^2x}{dt^2} = -6\pi\eta r\frac{dx}{dt} + F_r, \tag{3.10}$$

where m is the mass of the particle. Recall that,

$$\begin{aligned} \frac{1}{2}\frac{d(x)^2}{dt} &= \frac{dx}{dt}x \\ \frac{1}{2}\frac{d^2(x^2)}{dt^2} &= \frac{d^2x}{dt^2}x + \left(\frac{dx}{dt}\right)^2. \end{aligned} \tag{3.11}$$

Multiplying Equation (3.10) with x, substituting the identities (3.11), and taking expectations

$$\frac{m}{2}E\left[\frac{d^2(x^2)}{dt^2}\right] - mE\left[\left(\frac{dx}{dt}\right)^2\right] = -3\pi\eta rE\left[\frac{d(x)^2}{dt}\right] + E[F_rx]. \tag{3.12}$$

The relation between the average kinetic energy and temperature is

$$mE\left[\left(\frac{dx}{dt}\right)^2\right] = \frac{RT}{N}. \tag{3.13}$$

Assuming that the random force and the position are uncorrelated, $E[F_rx] = 0$ and defining a new variable, $z_1 = dE[x^2]/dt$, the following differential equation is produced,

$$\frac{m}{2}\frac{dz_1}{dt} - \frac{RT}{N} = -3\pi\eta rz_1 \tag{3.14}$$

which has the general solution

$$z_1(t) = \frac{RT}{N}\frac{1}{3\pi\eta r}[1 - exp\left(\frac{6\pi\eta r}{m}t\right)] \tag{3.15}$$

The exponential on the right goes to zero very quickly and therefore, the resulting mean squared displacement is nominally the constant multiplied with time,

$$z(t) = \frac{RT}{N}\frac{1}{3\pi\eta r}t. \tag{3.16}$$

Equation (3.16) is also the same equation derived by Einstein, Equation (3.11). The above model of Brownian motion is not seen as a solution to the white noise driven differential equation

$$\frac{d\beta(t)}{dt} = \omega(t), \tag{3.17}$$

but as the solution to the equation of the form

$$\frac{d^2\tilde{\beta}(t)}{dt^2} = -c\frac{d\tilde{\beta}(t)}{dt} + \omega(t) \tag{3.18}$$

within the limit of vanishing time constant. Langevin's version is called the physical Brownian motion and Einstein's version is called the mathematical Brownian motion. In this thesis, Brownian motion means the mathematical Brownian motion.

Consider a Brownian motion as a solution to Equation (3.17), where $\omega(t)$ is a white random process. "White" means that every value of $\omega(t)$ and $\omega(t^{'})$ are independent whenever $t \neq (t^{'})$.

Before moving on to the solution of SDEs, consider the derivative of Wiener Process(Brownian Motion), called the Gaussian White Noise in the following section.

3.2 Linear Differential Equations with Driving White Noise

A stochastic differential equation of the form

$$\frac{dx}{dt} = f(x,t) + L(x,t)\omega(t) \tag{3.19}$$

where $\omega(t)$ is some vector of forcing functions. Since the forcing function is random, the solution to the stochastic differential equation is a random process as well.

In the context of SDEs, the term $f(x,t)$ in Equation (3.19) is called the drift function which determines the nominal dynamics of the system, and $L(x,t)$ is the dispersion matrix which determines how the noise $\omega(t)$ enters the system.

The unknown function usually modeled as Gaussian and "white" in the sense $\omega(t)$ arises from the property that the power spectrum of white noise is constant (flat) over all frequencies. In mathematical sense white noise process can be defined as follows:

Definition 3.2.1. (White Noise). White noise process $\omega(t) \in \mathbb{R}^s$ is a random function with the following properties:
1. $\omega(t_1)$ and $\omega(t_2)$ are independent if $t_1 \neq t_2$.
2. $t \mapsto \omega(t)$ is a Gaussian process with zero mean and Dirac-delta-correlation :

$$\begin{aligned} m_w(t) &= E[\omega(t)] = 0 \\ C_w(t,s) &= E[\omega(t)\omega^T(s)] = \delta(t-s)Q, \end{aligned} \tag{3.20}$$

where Q is the spectral density of the process.

From the above properties we can also deduce the following somewhat peculiar properties of white noise:
1. The sample path $t \mapsto \omega(t)$ is discontinuous almost everywhere.
2. White noise is unbounded and it takes arbitrary large positive and negative values at any finite interval.

3.3 Heuristic Solutions of Linear SDEs

Consider a linear time-invariant stochastic differential equations (LTI SDEs) of the form:

$$\frac{dx(t)}{dt} = Fx(t) + Lw(t),\ x(0) \sim N(m_0, P_0), \tag{3.21}$$

where F and L are some constant matrices, white noise process $w(t)$ has zero mean and a given spectral density Q. In Equation (3.21), the specified random initial condition for the equation such that at initial time $t = 0$, the solutions should be Gaussian with a given mean m_0 and covariance P_0.

If assumed that the driving process $w(t)$ is deterministic and continuous, one can form the general solution to the differential equation as follows:

$$x(t) = exp(Ft)x(0) + \int_0^t exp(F(t-\tau))Lw(\tau)d\tau, \tag{3.22}$$

where $exp(Ft)$ is the matrix exponential function. Since the differential equation is linear, the solution is valid when $w(t)$ is a white noise process. The solution also turns out to be Gaussian, because the noise process is Gaussian and a linear differential equation can be considered as a linear operator acting on the noise process. White noise process has zero mean, taking expectations on both sides of Equation (3.22) yields,

$$E[x(t)] = exp(Ft)m_0, \tag{3.23}$$

which is the expected value of the SDE solutions over all realizations of noise. The mean function is here denoted as $m(t) = E[x(t)]$.

The covariance of the solution can be derived by substituting the solution into the definition of covariance and by using the delta-correlation property of white noise, which results in

$$\begin{aligned} E[(x(t) - m(t))(x(t) - m(t))^T] &= exp(Ft(P_0 exp(Ft)))^T \\ &+ \int_0^t exp(f(t-\tau))LQL^T exp(F(t-\tau))^T d\tau \end{aligned} \tag{3.24}$$

Denoting the covariance as $P(t) = E[x(t) - m(t))(x(t) - m)^T]$. Differentiating the mean and covariance solutions and collecting the terms we can also derive the following differential equations for the mean and covariance:

$$\begin{aligned} \frac{dm(t)}{dt} &= Fm(t) \\ \frac{dP(t)}{dt} &= FP(t) + P(t)F^T + LQL^T \end{aligned} \tag{3.25}$$

Despite the heuristic derivation, Equations (3.25) are indeed the correct differential equations for the mean and covariance. But it is easy to demonstrate that one has to be extremely careful in extrapolation of deterministic differential equation results to stochastic setting.
To derive the covariance differential equation, taking expectations on both sides of Equation (3.21)

$$E\left[\frac{dx(t)}{dt}\right] = E[Fx(t)] + E[Lw(t)], \tag{3.26}$$

Exchanging the order of expectation and differentiation, using the linearity of expectation and also recalling that white noise has zero mean then results in correct mean differential equation. One can attempt to do the same for the covariance. By chain rule of ordinary calculus produces,

$$\frac{d}{dt}\left[(x-m)(x-m)^T\right] = \left(\frac{dx}{dt} - \frac{dm}{dt}\right)(x-m)^T + (x-m)\left(\frac{dx}{dt} - \frac{dm}{dt}\right)^T, \tag{3.27}$$

Substituting the time derivatives to the right hand side and taking expectations,

$$\begin{aligned} \frac{d}{dt}E\left[(x-m)(x-m)^T\right] &= FE\left[(x(t) - m(t))(x(t) - m(t))^T\right] \\ &+ E\left[(x(t) - m(t))((x(t) - m(t))^T\right]F^T, \end{aligned} \tag{3.28}$$

which implies the covariance differential equation

$$\frac{dP(t)}{dt} = FP(t) + P(t)F^T. \tag{3.29}$$

But the equation is wrong, because the term $L(t)QL^T(t)$ is missing on the right hand side. The mistake is that we assume it is possible to use the product rule in Equation (3.27), which one cannot. This is one of the unique features of stochastic calculus and also shows us that we cannot just assume without analyzing solutions of SDEs via formal extensions of deterministic ODE solutions.

3.4 Heuristic Solutions of Non-Linear SDEs

Let us now analyze a differential equation of the form:

$$\frac{dx}{dt} = f(x,t) + L(x,t)w(t), \tag{3.30}$$

where $f(x,t)$ and $L(x,t)$ are non-linear functions and $w(t)$ is a white noise process with a spectral density Q. Unfortunately, one cannot take a similar approach to solve the equation using deterministic solutions like in the case of linear differential equations. The generalization of numerical methods for deterministic differential equations discussed in the previous chapter does not work since the basic requirement in those methods is the continuity of the right hand side and in fact, even differentiability of several orders. Because white noise is discontinuous everywhere, the right hand side is discontinuous everywhere and is certainly not differentiable anywhere either.

3.5 Existence and Uniqueness of Solutions

A solution to a stochastic differential equation is called strong, if for given Brownian motion $\beta(t)$ it is possible to construct a solution $s(t)$, which is unique for that given Brownian motion. It means that the whole path of the process is unique for a given Brownian motion. Hence strong uniqueness is also called path-wise uniqueness.

The strong uniqueness of a solution to SDE of the general form

$$dx = f(x,t)dt + L(x,t)d\beta,\ x(t_0) = x_0, \tag{3.31}$$

can be determined using stochastic Picard's iteration. Thus we first rewrite the equation in integral form

$$x(t) = x_0 + \int_{t_0}^{t} f(x(\tau), \tau) d\tau + \int_{t_0}^{t} L(x(\tau), \tau) d\beta(\tau). \tag{3.32}$$

Then the solution can be approximated with the following iteration. It can be shown

---

Stochastic Picard's iteration
Starting from the initial guess $\varphi_0(t) = x_0$. With the given β, computing approximations $\varphi_1(t), \varphi_2(t), \cdots$ via the following recursion:

$$\varphi_{n+1}(t) = x_0 + \int_{t_0}^{t} f(\varphi_n(\tau), \tau) d\tau + \int_{t_0}^{t} L(\varphi_n(\tau), \tau) d\beta(\tau). \tag{3.33}$$

This iteration converges to the exact solution in mean squared sense if both of the functions f and L grow at most linearly in x, and they are Lipschitz continuous in the same variable (see, e.g., Øksendal, 2003 [6]). If these conditions are met, then there exists a unique strong solution to the SDE.

A solution is called weak if it is possible to construct some Brownian motion $\tilde{\beta}(t)$ and a stochastic process $\tilde{x}(t)$ such that the pair is a solution to the stochastic differential equation. Weak uniqueness means that the probability law of the solution is unique, that is, there cannot be two solutions with different finite dimensional distributions. An existence of strong solution always implies the existence of a weak solution (every strong solution is also a weak solution), but the converse is not true. Determination if an equation has a unique weak solution when it does not have a unique strong solution is considerably harder than the criterion for the strong solution.

The next chapter discusses the probability distribution and the statistics of the SDEs and how the mean and covariance of linear SDEs are derived using the Fokker-Planck-Kolmogorov Equation.

Chapter 4

Statistics of SDEs [2]

4.1 Fokker-Planck-Kolmogorov Equation

In this section we derive the equation for the probability density of Itô process $x(t)$ (see Appendix A2), when the process is defined as the solution to the SDE

$$dx = f(x,t)dt + L(x,t)d\beta. \tag{4.1}$$

The probability density is usually denoted as $p(x(t))$, but in this section the density is actually a function of both x and t, thus we represent it as $p(x,t)$.

Theorem 4.1.1 (Fokker-Planck-Kolmogorov equation). *The probability density $p(x,t)$ of the solution of the SDE in equation (4.1) solves the partial differential equation*

$$\begin{aligned}\frac{\partial p(x,t)}{\partial t} &= -\sum_i \frac{\partial}{\partial x_i}[f_i(x,t)p(x,t)] \\ &\quad + \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}[L(x,t)QL^T(x,t)]_{ij}p(x,t).\end{aligned} \tag{4.2}$$

This partial differential equation is called the Fokker-Planck-Kolmogorov equation.

Proof. Let $\phi(x)$ be an arbitrary twice differential function. The Itô differential of $\phi(x(t))$

is, by the Itô formula(see A2), given as follows:

$$
\begin{aligned}
d\phi &= \sum_i \frac{\partial \phi}{\partial x_i} dx_i + \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 \phi}{\partial x_i \partial x_j} \right) dx_i dx_j \\
&= \sum_i \frac{\partial \phi}{\partial x_i} f_i(x,t) dt + \sum_i \frac{\partial \phi}{\partial x_i} [L(x,t) d\beta]_i \\
&+ \frac{1}{2} \sum_{ij} \left(\frac{\partial^2 \phi}{\partial x_i \partial x_j} [L(x,t) Q L^T(x,t)]_{ij} \right) dt.
\end{aligned} \tag{4.3}
$$

Taking expectations from both sides with respect to x and formally dividing by dt gives:

$$
\begin{aligned}
\frac{dE[\phi]}{dt} &= \sum_i E\left[\frac{d\phi}{dx_i} f_i(x,t) \right] \\
&+ \frac{1}{2} \sum_{ij} E\left[\left(\frac{\partial^2 \phi}{\partial x_i \partial x_j} \right) [L(x,t) Q L^T(x,t)]_{ij} \right].
\end{aligned} \tag{4.4}
$$

The left hand side of the equation can be written as

$$
\begin{aligned}
\frac{dE[\phi]}{dt} &= \frac{d}{dt} \int \phi(x) p(x,t) dx \\
&= \int \phi(x) \frac{\partial p(x,t)}{\partial t} dx.
\end{aligned} \tag{4.5}
$$

Using multidimensional integration by parts,

$$
\int_C \frac{\partial u(x)}{\partial x_i} \upsilon(x) dx = \int_\partial C u(x) \upsilon(x) n_i dS - \int_C u(x) \frac{\partial \upsilon(x)}{\partial x_i} dx, \tag{4.6}
$$

where n is the normal of the boundary ∂C of C and dS is it's area element. If the integration area is the whole $\mathbb{R}^\ltimes$ and the functions $u(x)$ and $\upsilon(x)$ vanish at infinity, then the boundary term on the right hand side vanishes, the formula becomes

$$
\int \frac{\partial u(x)}{\partial x_i} \upsilon(x) dx = - \int u(x) \frac{\partial \upsilon(x)}{dx_i} dx \tag{4.7}
$$

Consider the right hand side of Equation (4.3), the term inside the summation can be written as,

$$
\begin{aligned}
E\left[\frac{\partial \phi}{dx_i} f_i(x,t) \right] &= \int \frac{\partial \phi}{\partial x_i} f_i(x,t) p(x,t) dx \\
&= - \int \phi(x) \frac{\partial}{\partial x_i} [f_i(x,t) p(x,t)] dx,
\end{aligned} \tag{4.8}
$$

where by integration by parts with $u(x) = \phi(x)$ and $\upsilon(x) = f_i(x,t)p(x,t)$. For the term inside the summation of the second term yields,

$$
\begin{aligned}
E\left[\left(\frac{\partial^2 \phi}{\partial x_i \partial x_j}\right)[L(x,t)QL^T(x,t)]_{ij}\right] &= \int \left(\frac{\partial^2 \phi}{\partial x_i \partial x_j}\right)[L(x,t)QL^T(x,t)]_{ij}p(x,t)dx \\
&= -\int \left(\frac{\partial \phi}{\partial x_j}\right)\frac{\partial}{\partial x_i}\left[[L(x,t)QL^T(x,t)]_{ij}p(x,t)\right]dx \\
&= \int \phi(x)\frac{\partial^2}{\partial x_i \partial x_j}\left[[L(x,t)QL^T(x,t)]_{ij}p(x,t)\right]dx,
\end{aligned}
\tag{4.9}
$$

where we have first used the integration by parts formula with $u(x) = \partial\phi(x)/\partial x_j$, $\upsilon(x) = [L(x,t)QL^T(x,t)]_{ij}p(x,t)$ and then again with $u(x) = \phi(x)$, $\upsilon(x) = \frac{\partial}{\partial x_i}\left(\left[L(x,t)QL^T(x,t)\right]_{ij}p(x,t)\right)$.

If substituting Equations (4.7) (4.8) (4.9) into (4.1) we get:

$$
\begin{aligned}
\int \phi(x)\frac{\partial p(x,t)}{\partial t}dx = &-\sum_i \int \phi(x)\frac{\partial}{\partial x_i}[f_i(x,t)p(x,t)]dx \\
&+\frac{1}{2}\sum_{ij}\int \phi(x)\frac{\partial^2}{\partial x_i \partial x_j}[L(x,t)QL^T(x,t))]_{ij}p(x,t)dx,
\end{aligned}
\tag{4.10}
$$

which can be written as

$$
\begin{aligned}
&\int \phi(x)\left[\frac{\partial p(x,t)}{\partial t} + \sum_i \frac{\partial}{\partial x_i}[f_i(x,t)p(x,t)\right] \\
&-\frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial x_i \partial x_j}[L(x,t)QL^T(x,t)]_{ij}p(x,t)]dx = 0.
\end{aligned}
\tag{4.11}
$$

The only way the equation can be true for an arbitrary $\phi(x)$ is that the term within the brackets vanishes, which gives the FPK equation. □

4.2 Mean and Covariance of SDE

In the previous section we derived the Fokker-Planck-Kolmogorov (FPK) equation which, in principle, is the complete probabilistic description of the state. The mean, covariance and other moments of the state distribution can be derived from it's solution. However, we are often interested primarily on the mean and covariance of the distribution and would like to avoid solving the FPK equation as an intermediate step.

If we take a look at Equation (4.1), we see that it can be interpreted as equation for the general moments of the state distribution. The equation can be generalized to time dependent $\phi(x,t)$ by including the time derivative:

$$\begin{aligned}\frac{dE[\phi]}{dt} &= E\left[\frac{\partial \phi}{\partial t}\right] + \sum_i E\left[\frac{\partial \phi}{\partial x_i} f_i(x,t)\right] \\ &+ \frac{1}{2}\sum_{ij} E\left(\frac{\partial^2 \phi}{\partial x_i \partial x_j}\right)[L(x,t)QL^T(x,t)]_{ij}.\end{aligned} \tag{4.12}$$

If we select the function as $\phi(x,t) = x_u$, then the Equation (4.12) reduces to

$$\frac{dE[x_u]}{dt} = E[f_u(x,t)], \tag{4.13}$$

which can be seen as the differential equation for the components of the mean of the state. Let us denote the mean function as $m(t) = E[x(t)]$ and select the function as $\phi(x,t) = x_u x_v - m_u(t)m_v(t)$, then the equation gives

$$\begin{aligned}\frac{dE[x_u x_v - m_u(t)m_v(t)]}{dt} &= E[(x_v - m_v(t))f_u(x,t)] \\ &+ E[(x_u - m_u(t))f_v(x,t)] + [L(x,t)QL^T(x,t)]_{uv}.\end{aligned} \tag{4.14}$$

If we denote the covariance as $P(t) = E[x(t) - m(t))(x(t) - m(t))^T]$, then equations (4.13) and (4.14) can be written in the following matrix form:

$$\frac{dm}{dt} = E[f(x,t)] \tag{4.15}$$

$$\begin{aligned}\frac{dP}{dt} &= E\left[f(x,t)(x| - m)^T\right] + E\left[(x-m)f^T(x,t)\right] \\ &+ E\left[L(x,t)QL^T(x,t)\right],\end{aligned} \tag{4.16}$$

which are the differential equations for mean and covariance of the state. These equations cannot be used practically because the expectations should be taken with respect to the actual distribution of the state, which is the solution to the FPK equation. Only in Gaussian case the first two moments actually characterize the solution. Even though in non-linear case we cannot use these equations as such, they provide a useful starting point for forming Gaussian approximations to SDEs.

4.3 Higher Order Moments of SDEs

It is possible to derive differential equations for the higher order moments of SDEs, but the required number of equations quickly becomes huge, because if the state dimension is n, the number of independent third moments is cubic n_3 in the number of state dimension, the number of fourth order moments is quartic n_4 and so on.

Let us consider the scalar SDE

$$dx = f(x)dt + L(x)d\beta \tag{4.17}$$

We know that the expectation of an arbitrary twice differentiable function $\phi(x)$ satisfies

$$\frac{dE[\phi(x)]}{dt} = E\left[\frac{\partial\phi(x)}{\partial x}f(x)\right] + \frac{q}{2}E\left[\frac{\partial^2\phi(x)}{\partial x_2}L^2(x)\right]. \tag{4.18}$$

If we apply this to $\phi(x) = x^n$, where $n \geq 2$, produces

$$\frac{dE[x^n]}{dt} = nE[x^{n-1}f(x,t)] + \frac{q}{2}n(n-1)E[x^{n-2}L^2(x)], \tag{4.19}$$

which, in principle, gives the equations for third order moments, fourth order moments and so on. It is also possible to derive similar differential equations for the central moments, cumulants and quasi-moments.

However, unless $f(x)$ and $L(x)$ are linear(or affine) functions, the equation for the nth order moment depends on the moments of higher order $> n$. Therefore, in order to compute these expectations we would need to integrate an infinite number of moment equations which is not practically possible. This problem can be solved by using moment closure methods which typically are based on replacing the higher order moments (or cumulants or quasi-moments) with suitable approximations.

In scalar case, it is possible to form a distribution which has the given set of moments or cumulants or quasi-moments, for example, as the maximum entropy distribution. Unfortunately, in multidimensional case the situation is much more complex.

4.4 Mean and covariance of linear SDEs

Consider a linear stochastic differential equation of the general form

$$dx = F(t)x(t)dt + u(t)dt + L(t)d\beta(t), \tag{4.20}$$

where the initial conditions are $x(t_0) \sim N(m_0, P_0), F(t)$ and $L(t)$ are matrix valued functions of time, $u(t)$ is a vector valued function of time and $\beta(t)$ is a Brownian motion with diffusion matrix Q.

The mean and covariance can be solved from the Equations (4.15) and (4.16) which in this cases is reduced to

$$\begin{aligned} \frac{dm(t)}{dt} &= F(t)m(t) + u(t) \\ \frac{dP(t)}{dt} &= F(t)P(t) + P(t)F^T(t) + L(t)QL^T(t), \end{aligned} \tag{4.21}$$

with the initial conditions $m_0(t_0) = m_0$ and $P(t_0) = P_0$. The general solutions to these differential equations are

$$\begin{aligned} m(t) &= \Psi(t, t_0)m(t_0) + \int_{t_0}^{t} \Psi(t, \tau)u(\tau)d\tau \\ P(t) &= \Psi(t, t_0)P(t_0)\Psi^T(t, t_0) \\ &\quad + \int_{t_0}^{t} \Psi(t, \tau)L(\tau)QL^T(\tau)\Psi^T(t, \tau)d\tau, \end{aligned} \tag{4.22}$$

which could also be obtained by computing the mean and covariance of the solution in Equation (A.33).

Because the solution is a linear transformation of the Brownian motion, which is a Gaussian process, the solution is also Gaussian

$$p(x, t) = N(x(t)|m(t), P(t)), \tag{4.23}$$

which can be verified by checking that this distribution indeed solves the corresponding FPK equation (3.22).

In case of LTI SDE

$$dx = Fx(t)dt + Ld\beta(t), \tag{4.24}$$

the mean and covariance are also given by Equations (4.22). The only difference is that the matrices F, L as well as the diffusion of the Brownian motion Q are constant. In this LTI SDE, the transition matrix is the matrix exponential function $\Psi(t, \tau) =$

$exp(F(t-\tau))$ and the solutions to the differential equations reduce to

$$\begin{aligned} m(t) &= exp(F(t-t_0))m(t_0) \\ P(t) &= exp(F(t-t_0))P(t_0)exp(F(t-t_0))^T \\ &\quad + \int_{t_0}^{t} exp(F(t-\tau))LQL^T exp(F(t-\tau))^T d\tau, \end{aligned} \tag{4.25}$$

The covariance above can also be solved using matrix fractions (see e.g. [7]). If we define matrices $C(t)$ and $D(t)$ such that $P(t) = C(t)D^{-1}(t)$, it is easy to show that P solves the matrix Riccati differential equation

$$\frac{dP(t)}{dt} = FP(t) + P(t)F^T + LQL^T \tag{4.26}$$

if the matrices $C(t)$ and $D(t)$ solve the differential equation

$$\begin{pmatrix} dC(t)/dt \\ dD(t)/dt \end{pmatrix} = \begin{pmatrix} F & LQL^T \\ 0 & -F^T \end{pmatrix} \begin{pmatrix} C(t) \\ D(t) \end{pmatrix}, \tag{4.27}$$

and $P(t_0) = C(t_0)D(t_0)^{-1}$. Let us take an example,

$$\begin{aligned} C(t_0) &= P(t_0) \\ D(t_0) &= I \end{aligned} \tag{4.28}$$

Since the differential equation is linear and time invariant, it can be solved using the matrix exponential function:

$$\begin{pmatrix} C(t) \\ D(t) \end{pmatrix} = exp\left\{ \begin{pmatrix} F & LQL^T \\ 0 & -F^T \end{pmatrix} t \right\} \begin{pmatrix} C(t_0) \\ D(t_0) \end{pmatrix}. \tag{4.29}$$

The final solution is then given as $P(t) = C(t)D^{-1}(t)$. It is useful since both the mean and covariance can now be solved via simple matrix exponential function computation.

Now that we have a better understanding of the probability distribution and statistics of SDEs, in the next chapter we discuss the Numerical Methods that can be used to solve SDEs.

Chapter 5

Numerical Methods for Solving SDEs

This chapter discusses the algorithms for solutions of scalar SDEs and multivariate SDEs. The following are the methods:
1. Euler- Maruyamma Method.
2. Milstein's Method.
3. Itô-Taylor Method.
Firstly, the Itô-Taylor series for Stochastic Differential Equations is discussed briefly in the next section.

5.1 Itô-Taylor series of SDEs [2]

The Itô-Taylor series is an extension of the Taylor series of ODEs to SDEs (A1). The derivation is similar to the Taylor series. The Taylor series expansion of ODE consists of time derivatives of an input function. When these derivatives are replaced with the Itô formula, it forms the Itô-Taylor series of SDEs.

Consider the following SDE

$$dx = f(x(t), t)dt + L(x(t), t)d\beta \tag{5.1}$$

with the initial condition $x(t_0)$ and a probability distribution $p(x(t_0))$. $f(x(t), t)$ is a differentiable function, and $L(x(t), t)$ is the dispersion matrix which determines how the noise enters the system. β is a random variable. In case of multivariate SDEs, the vectors $x(t) \in \mathbb{R}^p$, $\beta \in \mathbb{R}^q$ and $f(x(t), t) \in \mathbb{R}^p$, while the matrix $L(x(t), t) \in \mathbb{R}^{p\times q}$. In integral form, the equation (5.2) can be expressed as

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x(\tau),\tau)\, d\tau + \int_{t_0}^{t} L(x(\tau),\tau)\, d\beta(\tau) \tag{5.2}$$

Applying the Itô formula (A.32) to the differentiation of the terms $f(x(t),t)$ and $L(x(t),t)$ produces,

$$\begin{aligned} df(x(t),t) &= \frac{\partial f(x(t),t)}{\partial t}dt + \sum_u \frac{\partial f(x(t),t)}{\partial x_u} f_u(x(t),t)dt \\ &+ \sum_u \frac{\partial f(x(t),t)}{\partial x_u}[L(x(t),t)d\beta(\tau)]_u \\ &+ \frac{1}{2}\sum_{uv} \frac{\partial^2 f(x(t),t)}{\partial x_u \partial x_v}[L(x(t),t)QL^T(x(t),t)]_{uv}dt, \end{aligned} \tag{5.3}$$

$$\begin{aligned} dL(x(t),t) &= \frac{\partial L(x(t),t)}{\partial t}dt + \sum_u \frac{\partial L(x(t),t)}{\partial x_u} f_u(x(t),t)dt \\ &+ \sum_u \frac{\partial L(x(t),t)}{\partial x_u}[L(x(t),t)d\beta(\tau)]_u \\ &+ \frac{1}{2}\sum_{uv} \frac{\partial^2 L(x(t),t)}{\partial x_u \partial x_v}[L(x(t),t)QL^T(x(t),t)]_{uv}dt. \end{aligned} \tag{5.4}$$

where Q is the diffusion of the Brownian motion.
Integrating the equations (5.3) and (5.4) across the time interval $[t_0, t]$ yields

$$\begin{aligned} f(x(t),t) &= f(x(t_0),t_0) + \int_{t_0}^{t} \frac{\partial f(x(\tau),\tau)}{\partial t}d\tau \\ &+ \int_{t_0}^{t} \sum_u \frac{\partial f(x(\tau),\tau)}{\partial x_u} f_u(x(\tau),\tau)d\tau \\ &+ \int_{t_0}^{t} \sum_u \frac{\partial f(x(\tau),\tau)}{\partial x_u}[L(x(\tau),\tau)\, d\beta(\tau)]_u \\ &+ \int_{t_0}^{t} \frac{1}{2}\sum_{uv} \frac{\partial^2 f(x(\tau),\tau)}{\partial x_u \partial x_v}[L(x(\tau),\tau)QL^T(x(\tau),\tau)]_{uv}d\tau, \end{aligned} \tag{5.5}$$

$$
\begin{aligned}
L(x(t),t) = L(x(t_0),t_0) &+ \int_{t_0}^{t} \frac{\partial L(x(\tau),\tau)}{\partial t} d\tau \\
&+ \int_{t_0}^{t} \sum_u \frac{\partial L(x(\tau),\tau)}{\partial x_u} f_u(x(\tau),\tau) d\tau \\
&+ \int_{t_0}^{t} \sum_u \frac{\partial L(x(\tau),\tau)}{\partial x_u} [L(x(\tau),\tau)\, d\beta(\tau)]_u \\
&+ \int_{t_0}^{t} \frac{1}{2} \sum_{uv} \frac{\partial^2 L(x(\tau),\tau)}{\partial x_u \partial x_v} [L(x(\tau),\tau) Q L^T(x(\tau),\tau)]_{uv} d\tau.
\end{aligned}
\tag{5.6}
$$

It is useful to define the following operators acting on a differentiable function $g \in \mathbb{R}^{p\times q}$,

$$
\begin{aligned}
\mathscr{L}_t g &= \frac{\partial g}{dt} + \sum_u \frac{\partial g}{\partial x_u} f_u + \frac{1}{2} \sum_{uv} \frac{\partial^2 g}{\partial x_u \partial x_v} [LQL^T]_{uv} \\
\mathscr{L}_{\beta,v} g &= \sum_u \frac{\partial g}{\partial x_u} L_{uv}, \quad u,v = 1,\cdots,p.
\end{aligned}
\tag{5.7}
$$

Using these defined operators, equations (5.5) and (5.6) can be written as,

$$
\begin{aligned}
f(x(t),t) = f(x(t_0),t_0) &+ \int_{t_0}^{t} \mathscr{L}_t f(x(\tau),\tau) d\tau \\
&+ \sum_v \int_{t_0}^{t} \mathscr{L}_{\beta,v} f(x(\tau),\tau) d\beta_v(\tau),
\end{aligned}
\tag{5.8}
$$

$$
\begin{aligned}
L(x(t),t) = L(x(t_0),t_0) &+ \int_{t_0}^{t} \mathscr{L}_t L(x(\tau),\tau) d\tau \\
&+ \sum_v \int_{t_0}^{t} \mathscr{L}_{\beta,v} L(x(\tau),\tau) d\beta_v(\tau).
\end{aligned}
\tag{5.9}
$$

Substituting the equations of (5.8) in equation (5.2) gives

$$
\begin{aligned}
x(t) = x(t_0) &+ f(x(t_0),t_0)(t-t_0) + L(x(t_0),t_0)(\beta(t) - \beta(t_0)) \\
&+ \int_{t_0}^{t} \int_{t_0}^{\tau} \mathscr{L}_t f(x(\tau).\tau) d\tau d\tau + \sum_v \int_{t_0}^{t} \int_{t_0}^{\tau} \mathscr{L}_{\beta,v} f(x(\tau),\tau) d\beta_v(\tau) d\tau \\
&+ \int_{t_0}^{t} \int_{t_0}^{\tau} \mathscr{L}_t L(x(\tau),\tau) d\tau d\beta(\tau) \\
&+ \sum_v \int_{t_0}^{t} \int_{t_0}^{\tau} \mathscr{L}_{\beta,v} L(x(\tau),\tau) d\beta_v(\tau) d\beta(\tau).
\end{aligned}
\tag{5.10}
$$

The equation (5.10) can be rewritten as

$$x(t) = x(t_0) + f(x(t_0), t_0)(t - t_0) + L(x(t_0), t_0)(\beta(t) - \beta(t_0)) + r_1(t), \tag{5.11}$$

where the remainder $r(t)$ is given by

$$\begin{aligned} r_1(t) = &\int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_t f(x(\tau), \tau) d\tau d\tau + \sum_v \int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_{\beta,v} f(x(t), t) d\beta_v(\tau) d\tau \\ &+ \int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_t L(x(\tau), \tau) d\tau d\beta(\tau) \\ &+ \sum_v \int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_{\beta,v} L(x(\tau), \tau) d\beta_v(\tau) d\beta(\tau). \end{aligned} \tag{5.12}$$

The first order approximation of equation (5.11) is given by,

$$x(t) \approx x(t_0) + f(x(t_0), t_0)(t - t_0) + L(x(t_0), t_0)(\beta(t) - \beta(t_0)) \tag{5.13}$$

This approximation is used to evaluate $x(t)$ in the algorithm Euler-Maruyama.

5.2 Euler-Maruyama Method [2]

In this method, $x(t)$ is approximated by simulating it as a discrete process denoted by $\hat{x}(t)$. For a SDE of the form (5.1), with initial condition $\hat{x}(t_0) = \hat{x}_0$ and a probability distribution $p(x_0)$, $\hat{x}(t)$ is simulated in K steps over the time interval $[0, t]$ with step size $\Delta t = t/K$. At each step k,

$$\hat{x}(t_{k+1}) = \hat{x}(t_k) + f(\hat{x}(t_k), t_k)\Delta t + L(\hat{x}(t_k), t_k)\Delta\beta_k \tag{5.14}$$

where $\Delta\beta_k$ is a random variable with the distribution $N(0, Q\Delta t)$ and $t_k = k\Delta t$, $k = 1, .., K$. In case of scalar SDEs, $N(0, Q\Delta t)$ is a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = Q\Delta t$. While for multivariate SDEs $N(0, Q\Delta t)$ is a normal distribution with zero mean and covariance $Q\Delta t$.

The Euler-Maruyama algorithm for scalar SDE in pseudo code is as follows

Algorithm 5.1 Euler-Maruyama method

$t = 0$, $\hat{x} = x_0$
for $k = 1$ **to** $t/\Delta t := K$ **do**
  $\Delta\beta \sim N(0, Q\Delta t)$
  $\hat{x}(t) = \hat{x}(t) + f(\hat{x}(t), t)\Delta t + L(\hat{x}(t), t)\Delta\beta$
end for

In case of multivariate SDEs, the algorithm for Euler-Maruyama method remains the same, except $x(t) \in \mathbb{R}^p$, $\beta \in \mathbb{R}^q$ and $f(x(t), t) \in \mathbb{R}^p$ are vectors, and $L(x(t), t) \in \mathbb{R}^{p \times q}$ is a matrix.

5.2.1 Properties of the Algorithm [2][8]

Definition 5.2.1. The strong order of convergence of a stochastic numerical integration method can be defined to be the smallest exponent γ, such that there exists a constant C that satisfies the inequality given below,

$$E[|x(t_k) - \hat{x}(t_k)|] \leq C\Delta t^\gamma \tag{5.15}$$

where E represents the expectation or mean, $x(t_k)$ is the actual value of $x(t)$ at $t_k = k\Delta t$, $\hat{x}(t_k)$ is the discretized approximation value of $x(t)$ at t_k and Δt is the step size.

Definition 5.2.2. The weak order of convergence can be defined to be the smallest exponent α, such that there exists a constant C that satisfies the inequality given below,

$$|E[g(x(t_n))] - E[g(\hat{x}(t_n))]| \leq C\Delta t^\alpha \tag{5.16}$$

for any function g.

It is shown in Kloeden and Platen 1999 [3]that the Euler-Maruyama method has strong order of convergence, $\gamma = 1/2$ and weak order of convergence, $\alpha = 1$. The strong order of convergence value is half of the weak order of convergence value due to the fact that the term $d\beta_v(\tau)d\beta(\tau)$ in the remainder (5.12) when integrated gives the term $d\beta(\tau)$ which is approximately equal to $dt^{1/2}$. Hence, expansion of this term can increase the strong order of convergence leading to the Milstein's approximation method for SDEs. h

5.3 Milstein's Method [2]

In equation (5.10), the term $\mathscr{L}_{\beta,v}L(x(\tau),\tau)$, can be expanded using the Itô-Taylor expansion like in equation (5.8) and (5.9), thus producing

$$\begin{aligned}\mathscr{L}_{\beta,v}L(x(t),t) &= \mathscr{L}_{\beta,v}L(x(t_0),t_0) + \int_{t_0}^{t} \mathscr{L}_t\mathscr{L}_{\beta,v}L(x(t),t)dt \\ &+ \sum_v \int_{t_0}^{t} \mathscr{L}_{\beta,v}^2 L(x(\tau),\tau)d\beta_v(\tau).\end{aligned} \tag{5.17}$$

Substituting equation (5.17) into equation (5.10) gives,

$$\begin{aligned}x(t) &= x(t_0) + f(x(t_0),t_0)(t-t_0) + L(x(t_0),t_0)(\beta(t)-\beta(t_0)) \\ &+ \sum_v \mathscr{L}_{\beta,v}L(x(t_0),t_0)\int_{t_0}^{t}\int_{t_0}^{\tau} d\beta_v(\tau)d\beta(\tau) + r_2(t)\end{aligned} \tag{5.18}$$

where the remainder $r_2(t)$ is given by

$$\begin{aligned}r_2(t) = &\int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_t f(x(\tau),\tau)d\tau d\tau + \sum_v \int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_{\beta,v} f(x(t),t)d\beta_v(\tau)d\tau \\ &+ \int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_t L(x(\tau),\tau)d\tau d\beta(\tau) + \sum_v \int_{t_0}^{t}\int_{t_0}^{\tau} \Bigg\{ \int_{t_0}^{t} \mathscr{L}_t\mathscr{L}_{\beta,v}L(x(t),t)dt \\ &+ \sum_v \int_{t_0}^{t} \mathscr{L}_{\beta,v}^2 L(x(\tau),\tau)d\beta_v(\tau)\Bigg\} d\beta_v(\tau)d\beta(\tau)\end{aligned} \tag{5.19}$$

Part of equation (5.18) evolves as a double iterated Itô integral given by

$$\int_{t_0}^{t}\int_{t_0}^{\tau} d\beta_v(\tau)d\beta(\tau). \tag{5.20}$$

This iterated Itô integral is quite complex to compute. However, assuming it can be computed and the corresponding Brownian increment can be drawn, the equation (5.18) can be approximated as follows

$$\begin{aligned}x(t) &\approx x(t_0) + f(x(t_0),t_0)(t-t_0) + L(x(t_0),t_0)(\beta(t)-\beta(t_0)) \\ &+ \sum_v \mathscr{L}_{\beta,v}L(x(t_0),t_0)\int_{t_0}^{t}\int_{t_0}^{\tau} d\beta_v(\tau)d\beta(\tau)\end{aligned} \tag{5.21}$$

This approximation is used to evaluate $x(t)$ in the algorithm of Milstein's method.

Similar to the Euler-Maruyama method, $x(t)$ is simulated as a discrete process given by $\hat{x}(t)$. For a SDE of the form (5.1), having initial condition $\hat{x}(t_0) = \hat{x}_0$ with a probability distribution $p(x_0)$, $\hat{x}(t)$ is simulated in K steps over the time interval $[0, t]$ with step size $\Delta t = t/K$. For multivariate system, $x(t) \in \mathbb{R}^p$, $\beta \in \mathbb{R}^q$ and $f(x(t), t) \in \mathbb{R}^p$ are vectors, and $L(x(t), t) \in \mathbb{R}^{p \times q}$ is a matrix. At each step k,

$$\begin{aligned} \hat{x}(t_{k+1}) &= \hat{x}(t_k) + f(\hat{x}(t_k), t_k)\Delta t + L(\hat{x}(t_k), t_k)\Delta\beta_k \\ &+ \sum_v \left[\sum_u \frac{\partial L}{\partial x_u}(\hat{x}(t_k), t_k) L_{uv}(\hat{x}(t_k), t_k) \right] \Delta\chi_{v,k} \end{aligned} \tag{5.22}$$

where $t_k = k\Delta t$, $k = 1, ...K$, $\Delta\beta_k = \beta(t_{k+1}) - \beta(t_k)$ is a Brownian motion increment and $\Delta\chi_{v,k}$ is it's related Itô integral given by

$$\Delta\chi_{v,k} = \int_{t_k}^{t_{k+1}} \int_{t_k}^{\tau} d\beta_v(\tau) d\beta(\tau) \tag{5.23}$$

The algorithm of Milstein's method for multivariate SDEs in pseudo code is given below,

Algorithm 5.2 Milstein's method for multivariate SDEs

$t = 0$, $\hat{x} = x_0$
for $k = 1$ **to** $t/\Delta t := K$ **do**
  $\Delta\beta = \beta(t_{k+1}) - \beta(t_k)$
  $\Delta\chi_v = \int_{k\Delta t}^{(k+1)\Delta t} \int_{k\Delta t}^{\tau} d\beta_v(\tau) d\beta(\tau)$
  $\hat{x}(t) = \hat{x}(t) + f(\hat{x}(t), t)\Delta t + L(\hat{x}(t), t)\Delta\beta$
  $+ \sum_v \left[\sum_u \frac{\partial L}{\partial x_u}(\hat{x}(t), t) L_{uv}(\hat{x}(t), t) \right] \Delta\chi_v$
end for

In case of scalar SDEs, the iterated Itô integral can be computed as:

$$\int_{t_0}^{t} \int_{t_0}^{\tau} d\beta\tau d\beta\tau = \frac{1}{2}[(\beta(t) - \beta(t_0))^2 - q(t - t_0)] \tag{5.24}$$

Hence, Milstein's method algorithm in pseudo code for scalar SDEs can be written explicitly as:

Algorithm 5.3 Scalar Milstein's method

$t = 0$, $\hat{x} = x_0$
for $k = 1$ **to** $t/\Delta t := K$ **do**
  $\Delta\beta \sim N(0, Q\Delta t)$
  $\hat{x}(t) = \hat{x}(t) + f(\hat{x}(t), t)\Delta t + L(\hat{x}(t), t)\Delta\beta + \frac{1}{2}\frac{\partial L}{\partial x}(\hat{x}(t), t) L(\hat{x}(t), t)(\Delta\beta_k^2 - Q\Delta t)$
end for

5.3.1 Properties of the Algorithm [2]

The strong and weak order of convergence of Milstein's method are both equal to 1. Considering that noise is additive, that is, $L(x, t) = L(t)$ the Milstein's algorithm reduces to Euler-Maruyama. Hence, in the case of additive noise, the strong order of convergence for Euler-Maruyama is 1.

Higher order Itô-Taylor series expansions can be formed by including more terms into the series. However, derivation of higher order methods involves drawing the iterated stochastic integral jointly with Brownian motion, which is difficult.

5.3.2 Example

A stochastic differential equation represented by Brownian motion,

$$\hat{x} = f(x(t), t)dt + \mathscr{L}(x(t), t)d\beta(t) \tag{5.25}$$

where $f(.)$ and $\mathscr{L}$ are the drift and diffusion coefficients and $\beta(t)$ is a Brownian motion. In the special case of Geometric Brownian Motion, where $f(.) = \mu\dot{x}$ and $\mathscr{L}(.) = \sigma\dot{x}$, the SDE is

$$\hat{x} = \mu x(t)dt + \sigma x(t)d\beta() \tag{5.26}$$

Rewriting the integral form of Geometric Brownian motion as follows:

$$x(t_{n+1}) - x(t_n) = \mu \int_{t_n}^{t_{n+1}} x(x)ds + \sigma \int_{t_n}^{t_{n+1}} x(s)d\beta s \tag{5.27}$$

The Euler-Maruyama approximation is

$$x_{n+1} - x_n = \mu x_n \Delta t_n + \sigma x_n \Delta\beta_n \tag{5.28}$$

The first integral is being approximated by $\mu x_n \delta t$ and the second integral by $\sigma x_n \delta\beta_n$. The Milstein method increases the accuracy of the E-M approximation with the addition of a second-order correction term derived from the stochastic Taylor series expansion[call appendix], yields the following differential form

$$x_{n+1} - x_n = f(x_n)\Delta t + (x_n)\Delta\beta_n + 0.5\sigma_2 x_n((\Delta\beta_N)^2 - \Delta t) \tag{5.29}$$

for the Geometric Brownian motion.

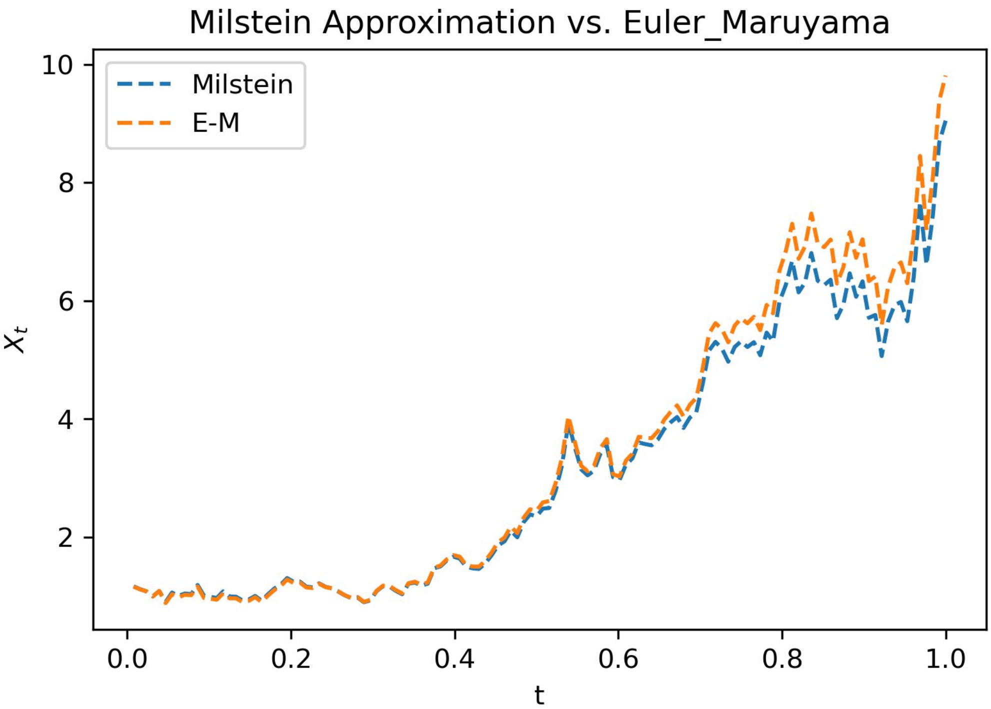

Figure 5.1 Milstein vs Euler-Maruyama

5.4 Itô-Taylor Method [2]

In a SDE of the form (5.1), consider L, the diffusion term denotes a constant matrix, that is $\mathscr{L}_t L = \mathscr{L}_{\beta,v} L = 0$. In addition to that, we assume the diffusion of Brownian motion Q as a constant. From (5.4), (5.5) and (5.7) the equation (5.10) reads,

$$\begin{aligned} x(t) &= x(t_0) + f(x(t_0), t_0)(t - t_0) + L(x(t_0), t_0)(\beta(t) - \beta(t_0)) \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_t f(x(\tau), \tau) d\tau d\tau + \sum_v \int_{t_0}^{t}\int_{t_0}^{\tau} \mathscr{L}_{\beta,v} f(x(t), t) d\beta_v d\tau \end{aligned} \tag{5.30}$$

Here, the composite terms $\mathscr{L}_t f(x(t),t)$ and $\mathscr{L}_{\beta,v} f(x(t),t)$ are expanded using the Itô-Taylor series (5.8) resulting in,

$$\begin{aligned} \mathscr{L}_t f(x(t),t) &= \mathscr{L}_t f(x(t_0),t_0) + \int_{t_0}^{t} \mathscr{L}_t^2 f(x(t),t)dt \\ &\quad + \sum_v \int_{t_0}^{t} \mathscr{L}_{\beta,v}\mathscr{L}_t f(x(t),t)d\beta_v \\ \mathscr{L}_{\beta,v} f(x(t),t) &= \mathscr{L}_{\beta,v} f(x(t_0),t_0) + \int_{t_0}^{t} \mathscr{L}_t\mathscr{L}_{\beta,v} f(x(t),t)dt \\ &\quad + \sum_v \int_{t_0}^{t} \mathscr{L}_{\beta,v}^2 f(x(t),t)d\beta_v \end{aligned} \tag{5.31}$$

Substituting equations of (5.31) in equation (5.30) gives

$$\begin{aligned} x(t) &= x(t_0) + f(x(t_0),t_0)(t-t_0) + L(x(t_0),t_0)(\beta(t)-\beta(t_0)) \\ &\quad + \mathscr{L}_t f(x(t_0),t_0)\frac{(t-t_0)^2}{2} \\ &\quad + \sum_v \mathscr{L}_{\beta,v} f(x(t_0),t_0)\int_{t_0}^{t}[\beta_v(\tau)-\beta_v(t_0)]d\tau + r_3(t), \end{aligned} \tag{5.32}$$

where the remainder $r_3(t)$ is given by

$$\begin{aligned} r_3(t) &= \int_{t_0}^{t}\int_{t_0}^{\tau}\left\{\int_{t_0}^{t}\mathscr{L}_t^2 f(x(t),t)dt + \sum_v \int_{t_0}^{t}\mathscr{L}_{\beta,v}\mathscr{L}_t f(x(t),t)d\beta_v\right\}d\tau d\tau \\ &\quad + \sum_v \int_{t_0}^{t}\int_{t_0}^{\tau}\left\{\int_{t_0}^{t}\mathscr{L}_t\mathscr{L}_{\beta,v} f(x(t),t)dt + \sum_v \int_{t_0}^{t}\mathscr{L}_{\beta,v}^2 f(x(t),t)d\beta_v\right\}d\beta_v(\tau)d\tau \end{aligned} \tag{5.33}$$

Thus, the equation (5.32) can be approximated as,

$$\begin{aligned} x(t) &\approx x(t_0) + f(x(t_0),t_0)(t-t_0) + L(x(t_0),t_0)(\beta(t)-\beta(t_0)) \\ &\quad + \mathscr{L}_t f(x(t_0),t_0)\frac{(t-t_0)^2}{2} + \sum_v \mathscr{L}_{\beta,v} f(x(t_0),t_0)\int_{t_0}^{t}[\beta_v(\tau)-\beta_v(t_0)]d\tau. \end{aligned} \tag{5.34}$$

The term $\beta(t)-\beta(t_0)$ and the integral $\int_{t_0}^{t}[\beta_v(\tau)-\beta_v(t_0)]d\tau$ refer to the same Brownian motion and thus are correlated. Both terms are Gaussian by assumptions, and so it follows that their joint distribution is:

$$\begin{bmatrix} \int_{t_0}^{t}[\beta(\tau)-\beta(t_0)] \\ \beta(t)-\beta(t_0) \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q(t-t_0)^3/3 & Q(t-t_0)^2/2 \\ Q(t-t_0)^2/2 & Q(t-t_0) \end{bmatrix} \right) \tag{5.35}$$

Thus, the equation (5.34) gives the Itô-Taylor method.

Because L and Q are constant, using the properties of the distribution (5.32), $\hat{s}$ is written in the form For a SDE of the form (5.1), with the given initial conditions and properties of the terms, $\hat{x}(t)$ is simulated in K steps over the time interval $[0, t]$ with step size $\Delta t = t/K$. At each step k, $\hat{x}(t_k)$ is computed as

$$\begin{aligned} \hat{x}(t_{k+1}) &= \hat{x}(t_k) + f(\hat{x}(t_k), t_k)\Delta t + L\Delta\beta_k \\ &\quad + a_k \frac{(\Delta t)^2}{2} + \sum_v b_{v,k}\Delta\zeta_k, \end{aligned} \tag{5.36}$$

where

$$\begin{aligned} a_k &= \frac{\partial f}{\partial t}(\hat{x}(t_k), t_k) + \sum_u \frac{\partial f}{\partial x_u}(\hat{x}(t_k), t_k) f_u(\hat{x}(t_k), t_k) \\ &\quad + \frac{1}{2}\sum_{uv} \frac{\partial^2 f}{\partial x_u \partial x_v}(\hat{x}(t_k), t_k)[LQL^T]_{uv} \\ b_{v,k} &= \sum_u \frac{\partial f}{\partial x_u}(\hat{x}(t_k), t_k) L_{uv}, \end{aligned} \tag{5.37}$$

and $\Delta\zeta_k$, $\Delta\beta_k$ are random variables defined using the joint distribution given by

$$\begin{bmatrix} \Delta\zeta_k \\ \Delta\beta_k \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q\Delta t^3/3 & Q\Delta^2/2 \\ Q\Delta t^2/2 & Q\Delta t \end{bmatrix} \right) \tag{5.38}$$

The algorithm for Itô-Taylor method in pseudo code is as follows:

Algorithm 5.4 Itô-Taylor Method

$t = 0$, $\hat{x} = x_0$

for $k = 1$ **to** $t/\Delta t := K$ **do**

$\quad \begin{bmatrix} \Delta\zeta \\ \Delta\beta \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q\Delta t^3/3 & Q\Delta^2/2 \\ Q\Delta t^2/2 & Q\Delta t \end{bmatrix} \right)$

$\quad a = \frac{\partial f}{\partial t}(\hat{x}(t), t) + \sum_u \frac{\partial f}{\partial x_u}(\hat{x}(t), t) f_u(\hat{x}(t), t) + \frac{1}{2}\sum_{uv} \frac{\partial^2 f}{\partial x_u \partial x_v}(\hat{x}(t), t)[LQL^T]_{uv}$

$\quad b_v = \sum_u \frac{\partial f}{\partial x_u}(\hat{x}(t), t) L_{uv}$

$\quad \hat{x}(t) = \hat{x}(t) + f(\hat{x}(t), t)\Delta t + L\Delta\beta + a\frac{(\Delta t)^2}{2} + \sum_v b_v \Delta\zeta$

end for

5.4.1 Properties of the Algorithm [2]

The strong order of convergence of the above Itô-Taylor method is shown to be 1.5 [3]. Hence, this method is also called as Strong order 1.5 Itô-Taylor method. This order is achieved by adding more terms from the Itô-Taylor expansion to Milstein's method.

When the SDE contains a diffusion term which is not constant, in both scalar and multivariate cases, one can use either the Euler-Maruyama Method or the Milstein's method, where Milstein's method has superior accuracy. The third Itô-Taylor method is a specialized one because it applies to only the special cases where L and the diffusion Q are constant. In the next chapter, an example power system [1] is considered and the numerical methods are used to simulate trajectories of discretization and diffusion.

Chapter 6

Application of the Numerical Methods for Solving SDEs

The methods are used to provide a good estimation of the dynamic state Jacobian matrix from the paper by Professor Wang [1]. In the paper, a proposed hybrid measurement and model based method is used to estimate the dynamic state Jacobian matrix, which provides invaluable information for various security analysis. Conventionally, the Jacobian matrix can be constructed based on state estimation results provided that an accurate dynamic model and network parameter values are available, which is not practically possible resulting in imprecise estimations. But, the proposed hybrid method does not depend on network model and can work as a robust alternative to the traditional state estimation based approaches when uncertainty in network topology is an issue.

6.1 Estimation of the Dynamic State Jacobian Matrix [1]

Considering the general power system dynamic model:

$$\begin{aligned} \dot{x} &= f(x, y) \\ O &= g(x, y) \end{aligned} \tag{6.1}$$

The above equations describe the generator dynamics (associated control) and the electrical transmission system and the static behavior of devices respectively. f and g are continuous functions, vectors $x \in \mathbb{R}^{n_x}$ and $y \in \mathbb{R}^{n_y}$ are the corresponding state variables (generator rotor angles, rotor speeds) and algebraic variables (bus voltages, bus angles).

In the paper, the proposed method uses the classical generator model, which can typically represents the generator dynamics in ambient conditions. Assuming that the load variations and renewable injections can be transformed into the variation on generator mechanical power, i.e., the mechanical power for Generator i is $P_{m_i} + \sigma_i x_i$, where x_i is the standard Gaussian noise, and σ_i^2 is the noise variance. Then, Equations (6.1) can be represented as:

$$\begin{aligned} \dot{\delta} &= \omega \\ M\dot{\omega} &= P_m - P_e - D\omega + \sum \xi \end{aligned} \tag{6.2}$$

with,

$$P_{e_i} = \sum_{j=i}^{3} E_i E_j (G_{ij} cos(\tilde{\delta}_i - \tilde{\delta}_j) + B_{ij} sin(\tilde{\delta}_i - \tilde{\delta}_j)) \tag{6.3}$$

Here, $\delta = [\delta_1, \cdots \delta_n]^T$ is a vector of generator rotor angles, $\omega = [\omega_1, \cdots, \omega_n]^T$ is a vector of generator rotor speeds, $P_m = [P_{m_1}, \cdots, P_{m_n}]^T$ is a vector of generators' mechanical power, $P_e = [P_{e_1}, \cdots, P_{e_n}]^T$ is a vector of generators' electrical power, $M = \mathrm{diag}(M_1, \cdots, M_n)$ are the inertia constants, $D = \mathrm{diag}(D_1, \cdots, D_n)$ are the damping factors. In addition, x_i is a vector of independent standard Gaussian random variables representing the variation of power injections, and $\sum = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$ is the covariance matrix, and constant impedances for triviality .

Linearizing Equations (6.2) gives,

$$\begin{bmatrix} \dot{\delta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} 0 & I_n \\ -M^{-1}\frac{\partial P_e}{\partial \delta} & -M^{-1}D \end{bmatrix} \begin{bmatrix} \delta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ M^{-1}\sum \end{bmatrix} \xi \tag{6.4}$$

Let $x = [\delta, \omega]^T, A = \begin{bmatrix} 0 & I_n \\ -M^{-1}\frac{\partial P_e}{\partial \delta} & -M^{-1}D \end{bmatrix}, B = [0, M^{-1}\sum]^T$, then Equation (6.4) becomes,

$$\dot{x} = Ax + B\xi \tag{6.5}$$

6.2 System Model [1]

Considering the standard WSCC 3-generator, 9-bus system model. The system model in the center-of-inertia (COI) formulation is given below:

$$\begin{aligned}
\dot{\tilde{\delta}}_1 &= \tilde{\omega}_1 \\
\dot{\tilde{\delta}}_2 &= \tilde{\omega}_2 \\
M_1\dot{\tilde{\omega}}_1 &= P_{m_1} - P_{e_1} - \frac{M_1}{M_T}P_{coi} - D_1\tilde{\omega}_1 + \sigma_1\xi_1 \\
M_2\dot{\tilde{\omega}}_2 &= P_{m_2} - P_{e_2} - \frac{M_2}{M_T}P_{coi} - D_2\tilde{\omega}_2 + \sigma_2\xi_2
\end{aligned} \tag{6.6}$$

where $\delta_0 = \frac{1}{M_T}\sum_{i=1}^{3} M_i\delta_i, \omega_0 = \frac{1}{M_T}\sum_{i=1}^{3} M_i\omega_i, M_T = \sum_{i=1}^{3} M_i, \tilde{\delta_i} = \delta_i - \delta_0, \tilde{\omega_i} = \omega_i - \omega_0$, for $i = 1, 2, 3$, and

$$P_{coi} = \sum_{i=1}^{3}(P_{m_i} - P_{e_i}) \tag{6.7}$$

Let the parameter values be assumed as: $P_{m_1} = 0.72\ p.u.$, $P_{m_2} = 1.63\ p.u.$, $P_{m_3} = 0.85\ p.u.$; $E_1 = 1.057 p.u.$, $E_2 = 1.050 p.u.$, $E_3 = 1.017 p.u.$; $M_1 = 0.63$, $M_2 = 0.34$, $M_3 = 0.16$; $D_1 = 0.63$, $D_2 = 0.34$, $D_3 = 0.16$.

Because the following relations that $\tilde{\delta_3} = -\frac{M_1\tilde{\delta_1}+M_2\tilde{\delta_2}}{M_3}$ and $\tilde{\omega_3} = -\frac{M_1\tilde{\omega_1}+M_2\tilde{\omega_2}}{M_3}$ hold in the COI formulation, $\tilde{\delta_3}$ and $\tilde{\omega_3}$ depending on the other state variables can be obtained without integration.

Thus, the system matrix can be written as follows:

$$A = \left[\begin{array}{cc|cc}
0 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{I}} \\
0 & 0 & & \\
\hline
\multicolumn{2}{c|}{\multirow{2}{*}{J}} & -\frac{D_1}{M_1} & 0 \\
 & & 0 & -\frac{D_2}{M_2}
\end{array}\right] \tag{6.8}$$

where $J = -M^{-1}\left(\frac{\partial P_e}{\partial \tilde{\delta}} + M\frac{1}{M_T}\frac{\partial P_{coi}}{\partial \tilde{\delta}}\right) = -M^{-1}\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi}$, for $i = 1, 2$;

$$\left(\left(\frac{\partial P_e}{\partial \tilde{\delta}}\right)_{coi}\right)_{ij} = \begin{cases} E_iE_j(G_{ij}sin(\tilde{\delta}_i - \tilde{\delta}_j) - B_{ij}cos(\tilde{\delta}_i - \tilde{\delta}_j)) & \\ +\frac{M_i}{M_T}\frac{\partial P_{coi}}{\partial \tilde{\delta}_i} & \text{if } i \neq j \\ \sum_{k=1}^{n} E_iE_k(G_{ik}sin(\tilde{\delta}_i - \tilde{\delta}_k) & \\ +B_{ik}cos(\tilde{\delta}_i - \tilde{\delta}_k)) + \frac{M_i}{M_T}\frac{\partial P_{coi}}{\partial \tilde{\delta}_i} & \text{if } i = j \end{cases} \tag{6.9}$$

where $\frac{\partial P_{coi}}{\partial \tilde{\delta}_i} = 2\sum_{k\neq i} E_iE_kG_{ik}sin(\tilde{\delta}_i - \tilde{\delta}_k)$. And the input matrix is as follows:

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_1 & 0 \\ 0 & 0 & 0 & \sigma_2 \end{bmatrix} \tag{6.10}$$

Thus, the system can be analysed using the numerical methods for SDE discussed in the previous chapter.

6.3 Derivatives of Itô- Taylor Method for the System model

Consider the initial conditions, $\delta_0 = 0$, $\omega_0 = 0$ and $\sigma_1 = \sigma_2 = 0.01$. Hence, $\tilde{\delta}_i = \delta_i$ and $\tilde{\omega}_i = \omega_i$. The system model discussed in section (6.2) can be written as

$$\dot{x} = Ax + B\xi \tag{6.11}$$

where

$$x = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \omega_1 \\ \omega_2 \end{bmatrix}; \quad A = \left[\begin{array}{c|c} \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & I \\ \hline J & \begin{matrix} -\dfrac{D_1}{M_1} & 0 \\ 0 & -\dfrac{D_2}{M_2} \end{matrix} \end{array}\right] \tag{6.12}$$

where $J = -M^{-1}\left(\frac{\partial P_e}{\partial \delta}\right)_{coi}$ for $i = 1, 2$;

$$\left(\left(\frac{\partial P_e}{\partial \delta}\right)_{coi}\right)_{ij} = \begin{cases} E_i E_j(G_{ij} sin(\delta_i - \delta_j) - B_{ij} cos(\delta_i - \delta_j)) \\ + \frac{M_i}{M_T}\frac{\partial P_{coi}}{\partial \delta_i} & \text{if } i \neq j \\ \sum_{k=1}^{n} E_i E_k(G_{ik} sin(\delta_i - \delta_k) + B_{ik} cos(\delta_i - \delta_k)) \\ + \frac{M_i}{M_T}\frac{\partial P_{coi}}{\partial \delta_i} & \text{if } i = j \end{cases} \tag{6.13}$$

where $\frac{\partial P_{coi}}{\partial \delta_i} = 2\sum_{k \neq i} E_i E_k G_{ik} sin(\delta_i - \delta_k)$. And the input matrix is as follows:

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & 0.01 \end{bmatrix} \tag{6.14}$$

and ξ is the noise input to the system. From equations (5.1) and (6.11), $f(x(t), t)$ and $L(x(t), t)$ can be written as,

$$f(x(t), t) = Ax \text{ and } L(x(t), t) = B \tag{6.15}$$

The Itô-Taylor algorithm for numerical solutions of SDE uses partial derivatives of $f(x(t), t)$ with respect to x and t.

Partial derivative of $f(x(t), t)$ with respect to t is given by

$$\frac{\partial}{\partial t}(f(x(t), t)) = \frac{\partial}{\partial t}(Ax) = \begin{bmatrix} \partial\omega_1/\partial t \\ \partial\omega_2/\partial t \\ \frac{\partial}{\partial t}(J_{11}\delta_1 + J_{12}\delta_2 - \frac{D_1}{M_1}\omega_1) \\ \frac{\partial}{\partial t}(J_{21}\delta_1 + J_{22}\delta_2 - \frac{D_2}{M_2}\omega_2) \end{bmatrix} \tag{6.16}$$

where J_{ij} is the element in i^{th} row and j^{th} column of matrix J. The terms of equation 6.16 are given by,

$$\begin{aligned}
\frac{\partial}{\partial t}\Big(J_{11}\delta_1 + J_{12}\delta_2 - \frac{D_1}{M_1}\omega_1\Big) = &-\frac{1}{M_1}\Bigg\{\sum_{k=1}^{3} E_1E_k\Bigg(\frac{\partial\delta_1}{\partial t}\big(G_{1k}sin(\delta_1-\delta_k) + B_{1k}cos(\delta_1-\delta_k)\big)\\
&+\delta_1\Big(\frac{\partial\delta_1}{\partial t} - \frac{\partial\delta_k}{\partial t}\Big)\big(G_{1k}cos(\delta_1-\delta_k) - B_{1k}sin(\delta_1-\delta_k)\big)\Bigg)\\
&+E_1E_2\Bigg(\big(G_{12}sin(\delta_1-\delta_2) - B_{12}cos(\delta_1-\delta_2)\big)\frac{\partial\delta_2}{\partial t}\\
&+\delta_2\Big(\frac{\partial\delta_1}{\partial t} - \frac{\partial\delta_2}{\partial t}\Big)\big(G_{12}cos(\delta_1-\delta_2) + B_{12}sin(\delta_1-\delta_2)\big)\Bigg)\\
&+2\,\frac{M_1}{M_T}\Bigg(\sum_{k=2,3} E_1E_k\Bigg(G_{1k}sin(\delta_1-\delta_k)\Big(\frac{\partial\delta_1}{\partial t} + \frac{\partial\delta_2}{\partial t}\Big)\\
&+(\delta_1+\delta_2)\Big(\frac{\partial\delta_1}{\partial t} - \frac{\partial\delta_k}{\partial t}\Big)G_{1k}cos(\delta_1-\delta_k)\Bigg)\Bigg)\Bigg\} - \frac{D_1}{M_1}\frac{\partial\omega_1}{\partial t}
\end{aligned} \tag{6.17}$$

and

$$\begin{aligned}
\frac{\partial}{\partial t}\Big(J_{21}\delta_1 + J_{22}\delta_2 - \frac{D_2}{M_2}\omega_2\Big) = &-\frac{1}{M_2}\Bigg\{\sum_{k=1}^{3} E_2E_k\Bigg(\frac{\partial\delta_2}{\partial t}\big(G_{2k}sin(\delta_2-\delta_k) + B_{2k}cos(\delta_2-\delta_k)\big)\\
&+\delta_2\Big(\frac{\partial\delta_2}{\partial t} - \frac{\partial\delta_k}{\partial t}\Big)\big(G_{2k}cos(\delta_2-\delta_k) - B_{2k}sin(\delta_2-\delta_k)\big)\Bigg)\\
&+E_2E_1\Bigg(\big(G_{21}sin(\delta_2-\delta_1) - B_{21}cos(\delta_2-\delta_1)\big)\frac{\partial\delta_1}{\partial t}\\
&+\delta_1\Big(\frac{\partial\delta_2}{\partial t} - \frac{\partial\delta_1}{\partial t}\Big)\big(G_{21}cos(\delta_2-\delta_1) + B_{21}sin(\delta_2-\delta_1)\big)\Bigg)\\
&+2\,\frac{M_2}{M_T}\Bigg(\sum_{k=1,3} E_2E_k\Bigg(G_{2k}sin(\delta_2-\delta_k)\Big(\frac{\partial\delta_1}{\partial t} + \frac{\partial\delta_2}{\partial t}\Big)\\
&+(\delta_1+\delta_2)\Big(\frac{\partial\delta_2}{\partial t} - \frac{\partial\delta_k}{\partial t}\Big)G_{2k}cos(\delta_2-\delta_k)\Bigg)\Bigg)\Bigg\} - \frac{D_2}{M_2}\frac{\partial\omega_2}{\partial t}
\end{aligned} \tag{6.18}$$

The partial derivative of $f(x(t),t) = Ax$ with respect to $x_1 = \delta_1$ is given by

$$\frac{\partial}{\partial \delta_1}(Ax) = \begin{bmatrix} 0 \\ 0 \\ \frac{\partial}{\partial \delta_1}(J_{11}\delta_1 + J_{12}\delta_2) \\ \frac{\partial}{\partial \delta_1}(J_{21}\delta_1 + J_{22}\delta_2) \end{bmatrix} \tag{6.19}$$

where

$$\begin{aligned} \frac{\partial}{\partial \delta_1}(J_{11}\delta_1 + J_{12}\delta_2) = & -\frac{1}{M_1}\Bigg\{ \sum_{k=2,3} E_1 E_k \Bigg(G_{1k}\big(sin(\delta_1 - \delta_k) + \delta_1 cos(\delta_1 - \delta_k)\big) \\ & + B_{1k}\big(cos(\delta_1 - \delta_k) - \delta_1 sin(\delta_1 - \delta_k)\big)\Bigg) + E_1^2 B_{11} \\ & + E_1 E_2 \delta_2 \big(G_{12} cos(\delta_1 - \delta_2) + B_{12} sin(\delta_1 - \delta_2)\big) \\ & + 2\,\frac{M_1}{M_T}\Big(\sum_{k=2,3} E_1 E_k G_{1k}\big(sin(\delta_1 - \delta_k) + (\delta_1 + \delta_2) cos(\delta_1 - \delta_k)\big)\Big)\Bigg\} \end{aligned} \tag{6.20}$$

and

$$\begin{aligned} \frac{\partial}{\partial \delta_1}(J_{21}\delta_1 + J_{22}\delta_2) = & -\frac{1}{M_2}\Bigg\{ \delta_2 E_2 E_1 \big(- G_{21} cos(\delta_2 - \delta_1) + B_{21} sin(\delta_2 - \delta_1)\big) \\ & + E_2 E_1 \Big(G_{21} sin(\delta_2 - \delta_1) - B_{21} cos(\delta_2 - \delta_1) - \delta_1 \big(G_{21} cos(\delta_2 - \delta_1) \\ & + B_{21} sin(\delta_2 - \delta_1)\big)\Big) + 2\,\frac{M_2}{M_T}\Big(E_2 E_1 G_{21}\big(sin(\delta_2 - \delta_1) \\ & - (\delta_1 + \delta_2) cos(\delta_2 - \delta_1)\big) + E_2 E_3 G_{23} sin(\delta_2 - \delta_3)\Big)\Bigg\} \end{aligned} \tag{6.21}$$

The partial derivative of $f(x(t),t) = Ax$ with respect to $x_2 = \delta_2$ is given by

$$\frac{\partial}{\partial \delta_2}(Ax) = \begin{bmatrix} 0 \\ 0 \\ \frac{\partial}{\partial \delta_2}(J_{11}\delta_1 + J_{12}\delta_2) \\ \frac{\partial}{\partial \delta_2}(J_{21}\delta_1 + J_{22}\delta_2) \end{bmatrix} \tag{6.22}$$

where

$$
\begin{aligned}
\frac{\partial}{\partial \delta_2}(J_{11}\delta_1 + J_{12}\delta_2) = &- \frac{1}{M_1}\Bigg\{ \delta_1 E_1 E_2\big(- G_{12}cos(\delta_1 - \delta_2) + B_{12}sin(\delta_1 - \delta_2)\big) \\
&+ E_1 E_2\Big(G_{12}sin(\delta_1 - \delta_2) - B_{12}cos(\delta_1 - \delta_2) - \delta_2\big(G_{12}cos(\delta_1 - \delta_2) \\
&+ B_{12}sin(\delta_1 - \delta_2)\big)\Big) + 2\,\frac{M_1}{M_T}\Big(E_1 E_2 G_{12}\big(sin(\delta_1 - \delta_2) \\
&- (\delta_1 + \delta_2)cos(\delta_1 - \delta_2)\big) + E_1 E_3 G_{13}sin(\delta_1 - \delta_3)\Big)\Bigg\}
\end{aligned}
\tag{6.23}
$$

and

$$
\begin{aligned}
\frac{\partial}{\partial \delta_2}(J_{21}\delta_1 + J_{22}\delta_2) = &- \frac{1}{M_2}\Bigg\{ \sum_{k=1,3} E_2 E_k\Big(G_{2k}\big(sin(\delta_2 - \delta_k) + \delta_2 cos(\delta_2 - \delta_k)\big) \\
&+ B_{2k}\big(cos(\delta_2 - \delta_k) - \delta_2 sin(\delta_2 - \delta_k)\big)\Big) + E_2^2 B_{22} \\
&+ E_2 E_1 \delta_1\big(G_{21}cos(\delta_2 - \delta_1) + B_{21}sin(\delta_2 - \delta_1)\big) \\
&+ 2\,\frac{M_2}{M_T}\Big(\sum_{k=1,3} E_2 E_k G_{2k}\big(sin(\delta_2 - \delta_k) + (\delta_1 + \delta_2)cos(\delta_2 - \delta_k)\big)\Big)\Bigg\}
\end{aligned}
\tag{6.24}
$$

The partial derivative of $f(x(t), t) = Ax$ with respect to $x_3 = \omega_1$ is given by

$$
\frac{\partial}{\partial \omega_1}(Ax) = \begin{bmatrix} 1 \\ 0 \\ \frac{-D_1}{M_1} \\ 0 \end{bmatrix} \tag{6.25}
$$

The partial derivative of $f(x(t), t) = Ax$ with respect to $x_4 = \omega_2$ is given by

$$
\frac{\partial}{\partial \omega_2}(Ax) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \frac{-D_2}{M_2} \end{bmatrix} \tag{6.26}
$$

The partial derivative of equation (6.19) with respect to $x_1 = \delta_1$ is given by

$$\frac{\partial^2}{\partial\delta_1^2}(Ax) = \begin{bmatrix} 0 \\ 0 \\ \frac{\partial^2}{\partial\delta_1^2}(J_{11}\delta_1 + J_{12}\delta_2) \\ \frac{\partial^2}{\partial\delta_1^2}(J_{21}\delta_1 + J_{22}\delta_2) \end{bmatrix} \tag{6.27}$$

where

$$\begin{aligned} \frac{\partial^2}{\partial\delta_1^2}(J_{11}\delta_1 + J_{12}\delta_2) = -\frac{1}{M_1}\Bigg\{ & \sum_{k=2,3} E_1E_k\Big(G_{1k}\big(2cos(\delta_1 - \delta_k) - \delta_1 sin(\delta_1 - \delta_k)\big) \\ & - B_{1k}\big(2sin(\delta_1 - \delta_k) + \delta_1 cos(\delta_1 - \delta_k)\big)\Big) \\ & + E_2E_1\delta_2\big(- G_{12}sin(\delta_1 - \delta_2) + B_{12}cos(\delta_1 - \delta_k)\big) \\ & + 2\,\frac{M_1}{M_T}\Big(\sum_{k=2,3} E_1E_kG_{1k}\big(2cos(\delta_1 - \delta_k) - (\delta_1 + \delta_2)sin(\delta_1 - \delta_k)\big)\Big)\Bigg\} \end{aligned} \tag{6.28}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial\delta_1^2}(J_{21}\delta_1 + J_{22}\delta_2) = -\,\frac{1}{M_2}\Bigg\{ & -\delta_2E_2E_1\big(G_{21}sin(\delta_2 - \delta_1) + B_{21}cos(\delta_2 - \delta_1)\big) \\ & - E_2E_1\Big(2\big(G_{21}cos(\delta_2 - \delta_1) + B_{21}sin(\delta_2 - \delta_1)\big) \\ & + \delta_1\big(G_{21}sin(\delta_2 - \delta_1) - B_{21}cos(\delta_2 - \delta_1)\big)\Big) \\ & - 2\,\frac{M_2}{M_T}E_2E_1G_{21}\big(2cos(\delta_2 - \delta_1) + (\delta_1 + \delta_2)sin(\delta_2 - \delta_1)\big)\Bigg\} \end{aligned} \tag{6.29}$$

The partial derivative of equation (6.22) with respect to $x_2 = \delta_2$ is given by

$$\frac{\partial^2}{\partial\delta_2^2}(Ax) = \begin{bmatrix} 0 \\ 0 \\ \frac{\partial^2}{\partial\delta_2^2}(J_{11}\delta_1 + J_{12}\delta_2) \\ \frac{\partial^2}{\partial\delta_2^2}(J_{21}\delta_1 + J_{22}\delta_2) \end{bmatrix} \tag{6.30}$$

where

$$
\begin{aligned}
\frac{\partial^2}{\partial \delta_2^2}(J_{11}\delta_1 + J_{12}\delta_2) = -\frac{1}{M_1}\Bigg\{ & -\delta_1 E_1 E_2 \big(G_{12}sin(\delta_1-\delta_2) + B_{12}cos(\delta_1-\delta_2)\big) \\
& - E_1E_2\Big(2\big(G_{12}cos(\delta_1-\delta_2) + B_{12}sin(\delta_1-\delta_2)\big) \\
& + \delta_2\big(G_{12}sin(\delta_1-\delta_2) - B_{12}cos(\delta_1-\delta_2)\big)\Big) \\
& - 2\,\frac{M_1}{M_T}E_1E_2G_{12}\big(2cos(\delta_1-\delta_2) + (\delta_1+\delta_2)sin(\delta_1-\delta_2)\big)\Bigg\}
\end{aligned}
\tag{6.31}
$$

and

$$
\begin{aligned}
\frac{\partial^2}{\partial \delta_2^2}(J_{12}\delta_1 + J_{22}\delta_2) = -\frac{1}{M_2}\Bigg\{ & \sum_{k=1,3} E_2E_k\Big(G_{2k}\big(2cos(\delta_2-\delta_k) - \delta_2 sin(\delta_2-\delta_k)\big) \\
& - B_{2k}\big(2sin(\delta_2-\delta_k) + \delta_2 cos(\delta_2-\delta_k)\big)\Big) \\
& + E_2E_1\delta_1\big(-G_{21}sin(\delta_2-\delta_1) + B_{21}cos(\delta_2-\delta_k)\big) \\
& + 2\,\frac{M_2}{M_T}\Big(\sum_{k=1,3} E_2E_kG_{2k}\big(2cos(\delta_2-\delta_k) - (\delta_1+\delta_2)sin(\delta_2-\delta_k)\big)\Big)\Bigg\}
\end{aligned}
\tag{6.32}
$$

The partial derivative of equation (6.25) with respect to $x_3 = \omega_1$ is given by

$$\frac{\partial^2}{\partial \omega_1^2}(Ax) = 0 \tag{6.33}$$

The partial derivative of equation (6.26) with respect to $x_4 = \omega_2$ is given by

$$\frac{\partial^2}{\partial \omega_2^2}(Ax) = 0 \tag{6.34}$$

The partial derivative of equation (6.19) with respect to $x_3 = \omega_1$ is given by

$$\frac{\partial^2(Ax)}{\partial \delta_1 \partial \omega_1} = \frac{\partial^2(Ax)}{\partial \omega_1 \partial \delta_1} = 0 \tag{6.35}$$

The partial derivative of equation (6.19) with respect to $x_4 = \omega_2$ is given by

$$\frac{\partial^2(Ax)}{\partial \delta_1 \partial \omega_2} = \frac{\partial^2(Ax)}{\partial \omega_2 \partial \delta_1} = 0 \tag{6.36}$$

The partial derivative of equation (6.22) with respect to $x_3 = \omega_1$ is given by

$$\frac{\partial^2(Ax)}{\partial\delta_2\partial\omega_1} = \frac{\partial^2(Ax)}{\partial\omega_1\partial\delta_2} = 0 \tag{6.37}$$

The partial derivative of equation (6.22) with respect to $x_4 = \omega_2$ is given by

$$\frac{\partial^2(Ax)}{\partial\delta_2\partial\omega_2} = \frac{\partial^2(Ax)}{\partial\omega_2\partial\delta_2} = 0 \tag{6.38}$$

The partial derivative of equation (6.25) with respect to $x_4 = \omega_2$ is given by

$$\frac{\partial^2(Ax)}{\partial\omega_1\partial\omega_2} = \frac{\partial^2(Ax)}{\partial\omega_2\partial\omega_1} = 0 \tag{6.39}$$

The partial derivative of equation (6.19) with respect to $x_2 = \delta_2$ is given by

$$\frac{\partial^2(Ax)}{\partial\delta_1\partial\delta_2} = \frac{\partial^2(Ax)}{\partial\delta_2\partial\delta_1} = \begin{bmatrix} 0 \\ 0 \\ \frac{\partial^2(J_{11}\delta_1+J_{12}\delta_2)}{\partial\delta_1\partial\delta_2} \\ \frac{\partial^2(J_{21}\delta_1+J_{22}\delta_2)}{\partial\delta_1\partial\delta_2} \end{bmatrix} \tag{6.40}$$

where

$$\begin{aligned} \frac{\partial^2(J_{11}\delta_1 + J_{12}\delta_2)}{\partial\delta_1\partial\delta_2} &= -\frac{1}{M_1}\Big\{E_1E_2\Big(G_{12}(\delta_1+\delta_2)sin(\delta_1-\delta_2) + B_{12}\big(2sin(\delta_1-\delta_2) \\ &\quad + (\delta_1-\delta_2)cos(\delta_1-\delta_2)\big)\Big) + 2\,\frac{M_1}{M_T}\Big(E_1E_2G_{12}(\delta_1+\delta_2)sin(\delta_1-\delta_2) \\ &\quad + E_1E_3G_{13}cos(\delta_1-\delta_2)\Big)\Big\} \end{aligned} \tag{6.41}$$

and

$$\begin{aligned} \frac{\partial^2(J_{21}\delta_1 + J_{22}\delta_2)}{\partial\delta_1\partial\delta_2} &= -\frac{1}{M_2}\Big\{E_2E_1\Big(G_{21}(\delta_1+\delta_2)sin(\delta_2-\delta_1) + B_{21}\big(2sin(\delta_2-\delta_1) \\ &\quad + (\delta_2-\delta_1)cos(\delta_2-\delta_1)\big)\Big) + 2\,\frac{M_2}{M_T}\Big(E_2E_1G_{21}(\delta_1+\delta_2)sin(\delta_2-\delta_1) \\ &\quad + E_2E_3G_{23}cos(\delta_2-\delta_3)\Big)\Big) \end{aligned} \tag{6.42}$$

These partial derivatives from equation (6.16) to equation (6.42) are used to compute the discrete process $\hat{x}$ in the algorithm for Itô-Taylor method.

6.4 Euler-Maruyama Method

The system model in section (6.3) is approximated using the algorithm for Euler-Maruyama method discussed in the section (5.2). The plots of trajectory of x with respect to t are as follows:

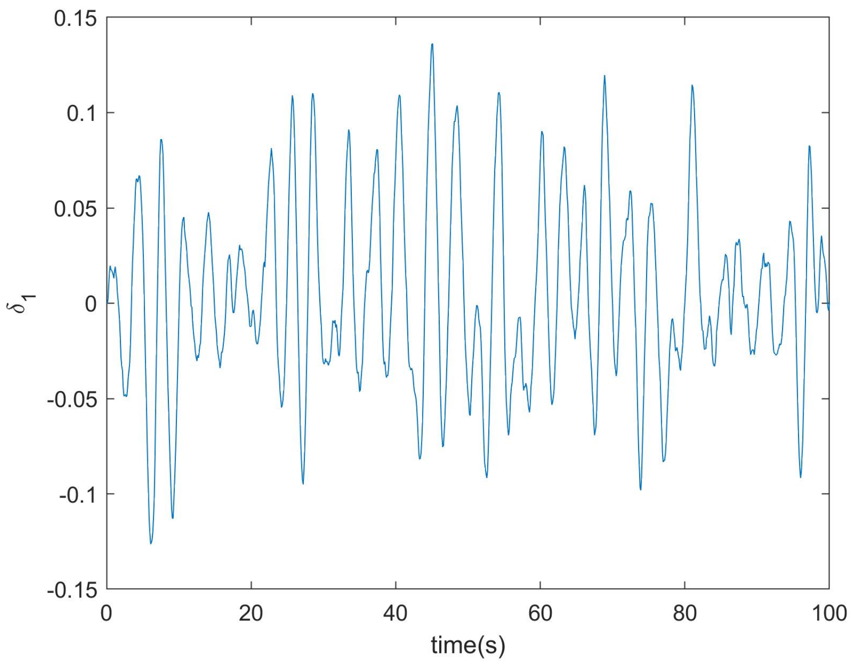

Figure 6.1 Trajectory of δ_1 on $[0, 100s]$

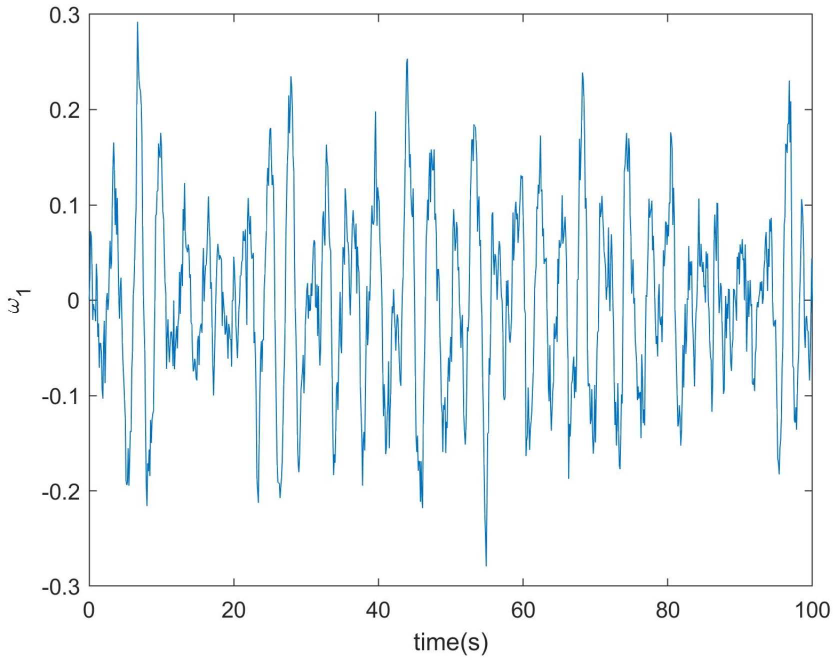

Figure 6.2 Trajectory of ω_1 on $[0, 100s]$

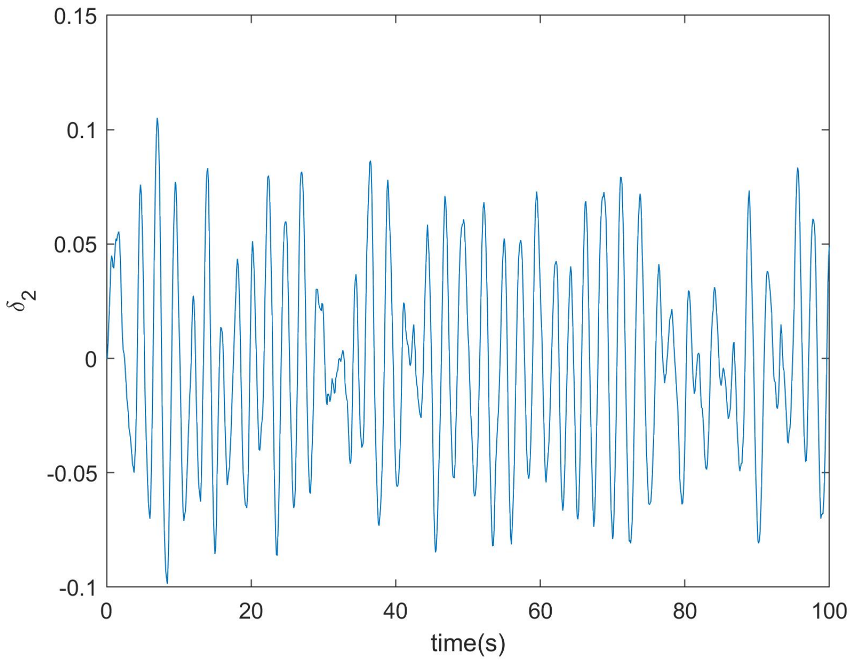

Figure 6.3 Trajectory of δ_2 on $[0, 100s]$

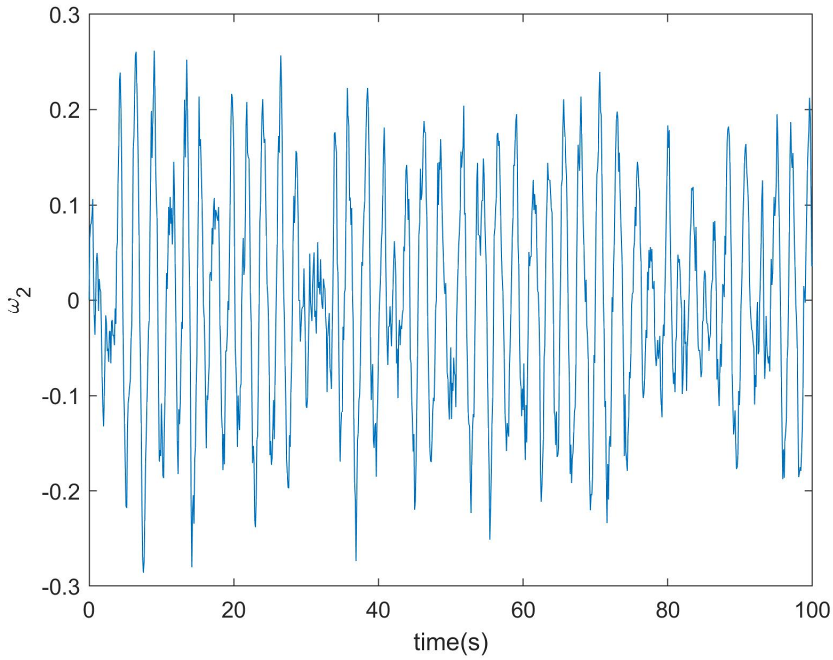

Figure 6.4 Trajectory of ω_2 on $[0, 100s]$

The system model in section (6.3) is approximated with varying Δt to analyse the behaviour of the algorithm with discretization. The plots of trajectories of x are as follows:

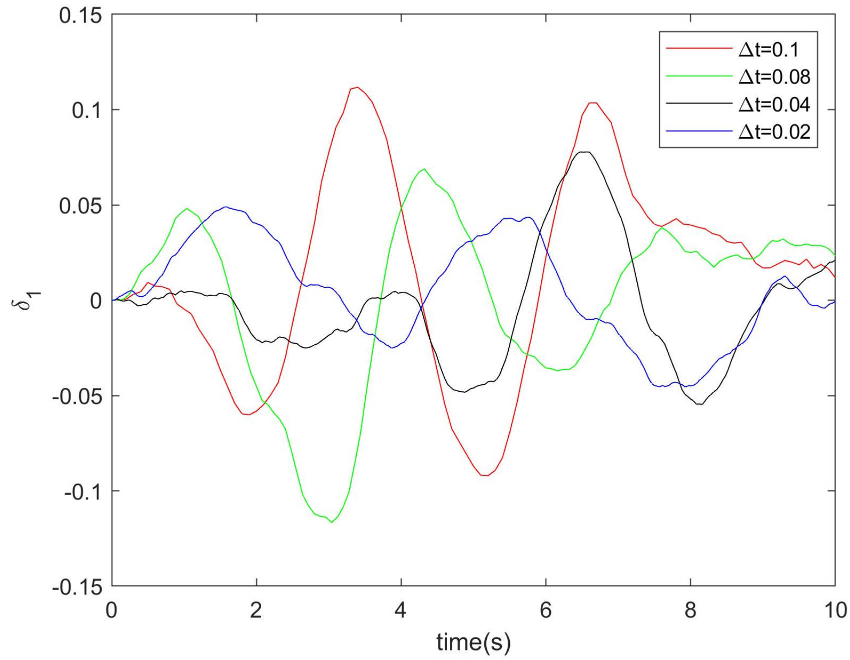

Figure 6.5 Trajectories of δ_1 for varying Δt

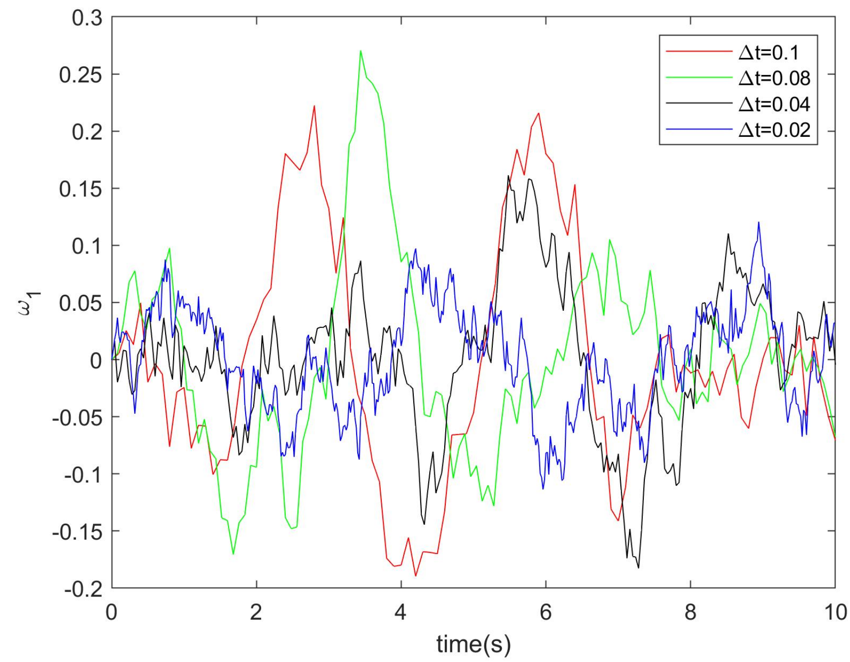

Figure 6.6 Trajectories of ω_1 for varying Δt

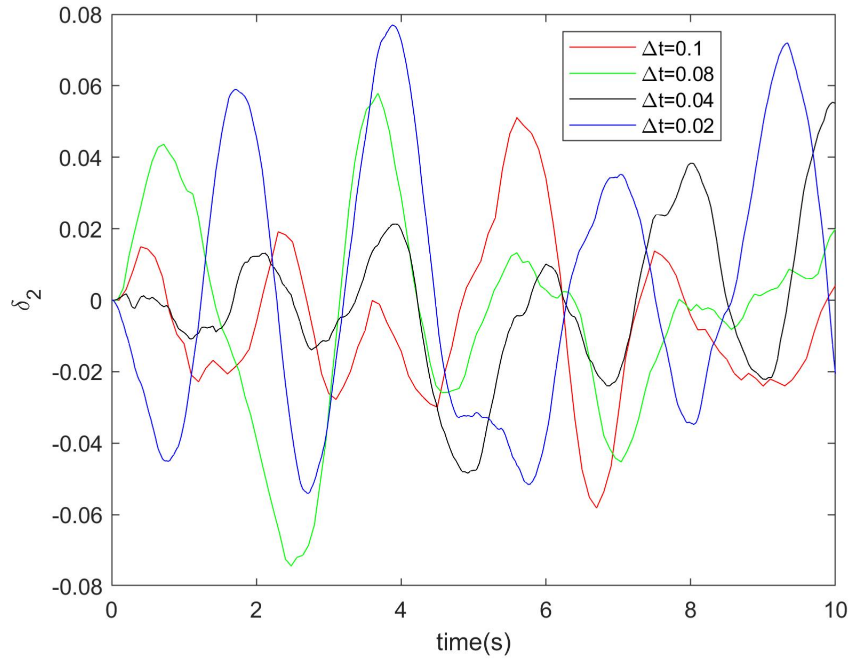

Figure 6.7 Trajectories of δ_2 for varying Δt

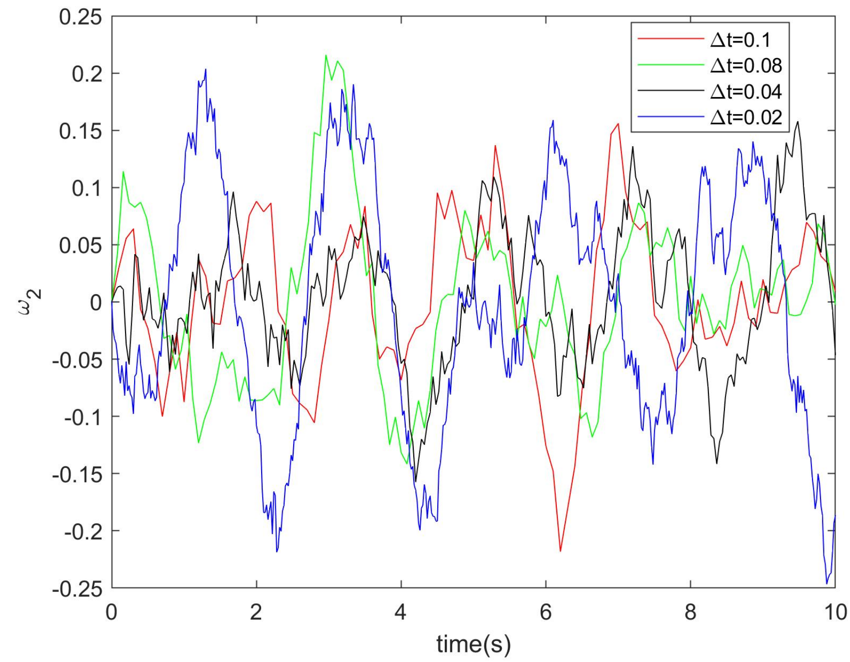

Figure 6.8 Trajectories of ω_2 for varying Δt

When the diffusion of the Brownian motion Q is varied and the system equations are approximated, the plots of trajectories of x are as follows:

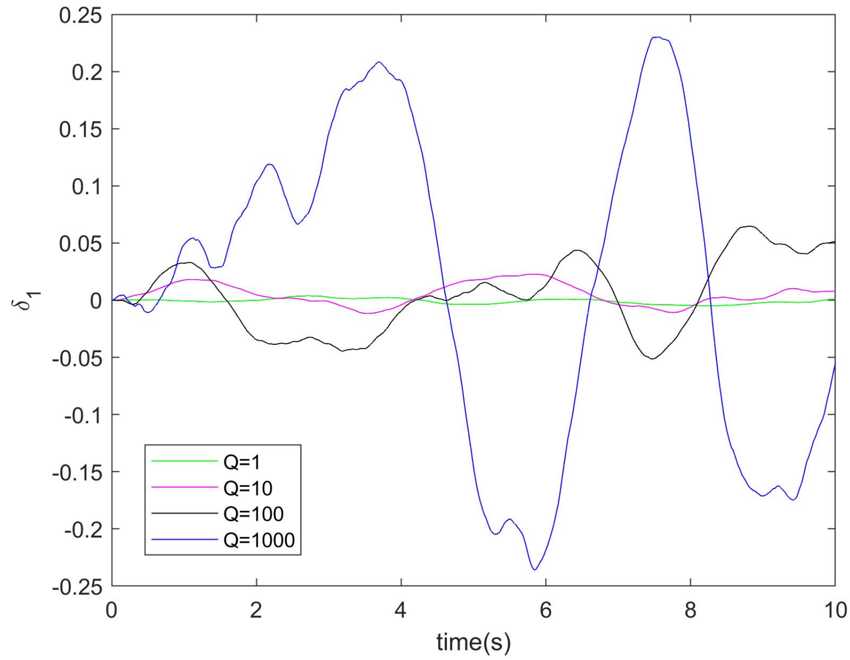

Figure 6.9 Trajectories of δ_1 for varying Q

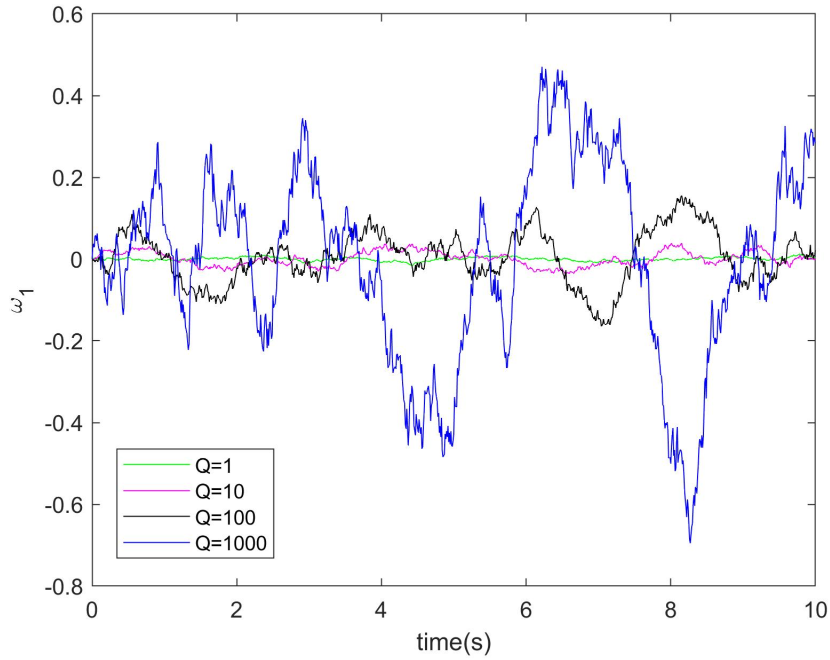

Figure 6.10 Trajectories of ω_1 for varying Q

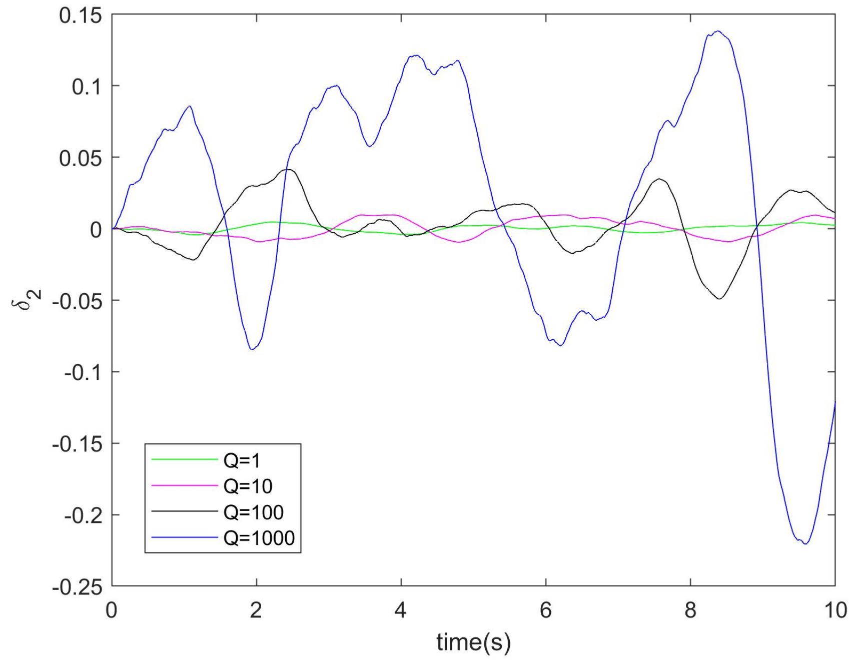

Figure 6.11 Trajectories of δ_2 for varying Q

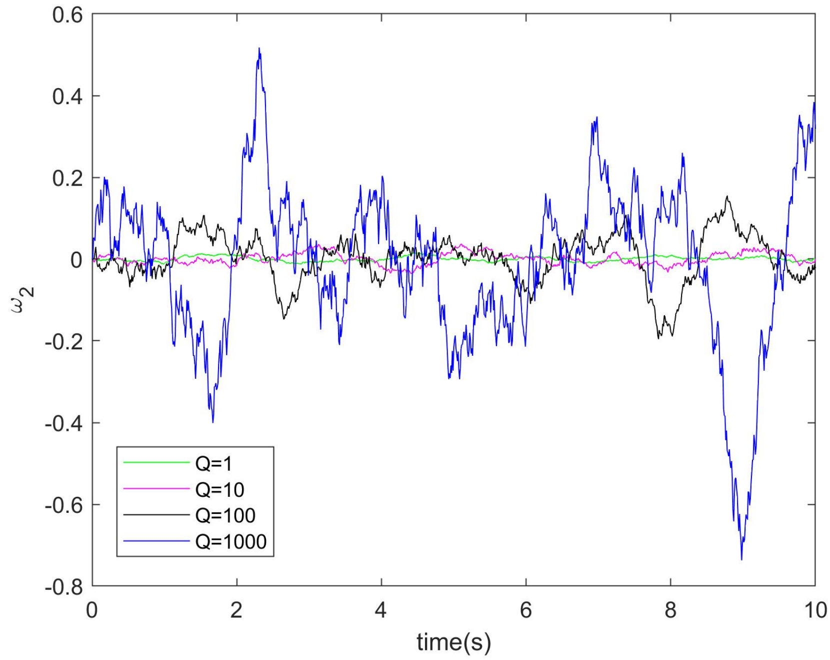

Figure 6.12 Trajectories of ω_2 for varying Q

6.5 Milstein's Method

The system equations in section (6.3) are approximated using the algorithm for Milstein's method discussed in section (5.3). The trajectory plots of x with respect to t are as follows:

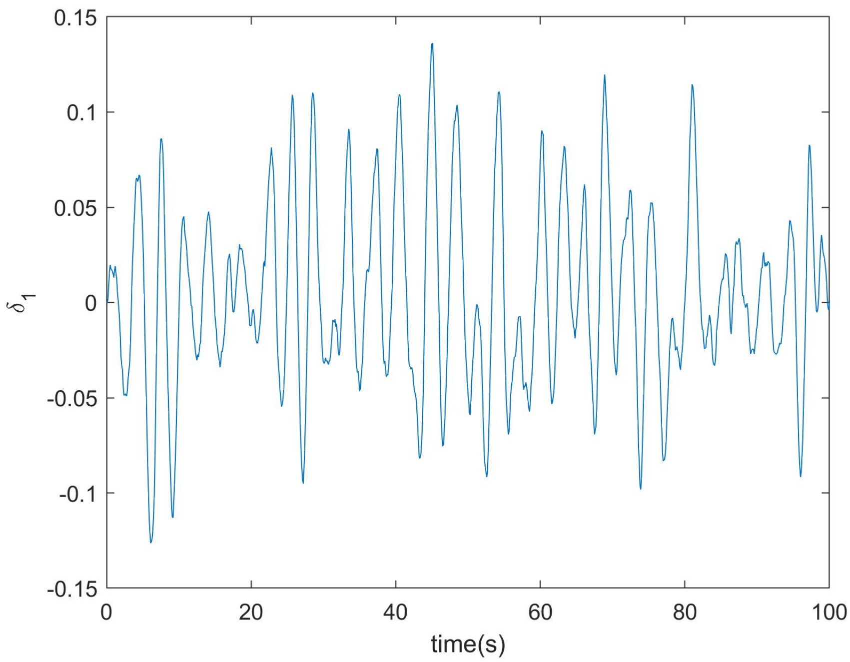

Figure 6.13 Trajectory of δ_1 on $[0, 100s]$

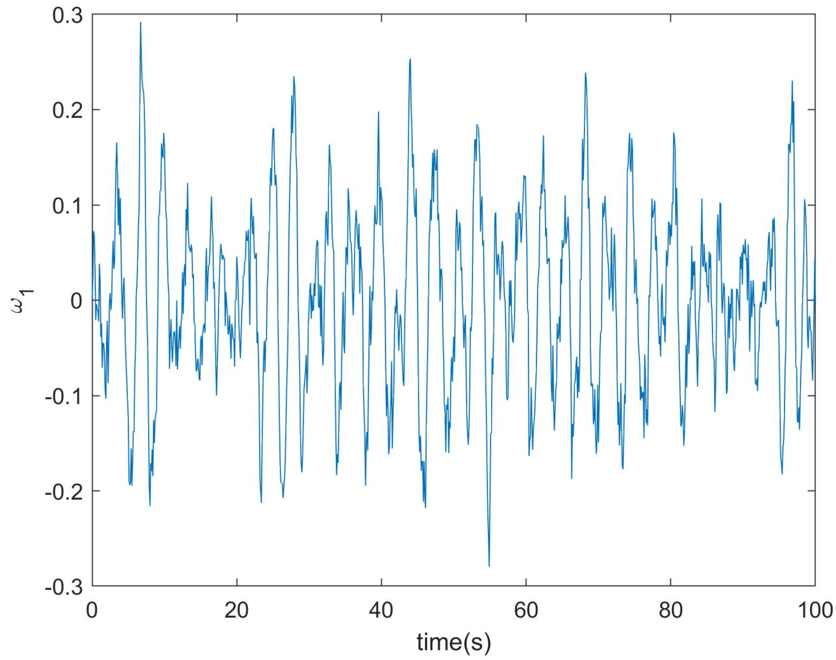

Figure 6.14 Trajectory of ω_1 on $[0, 100s]$

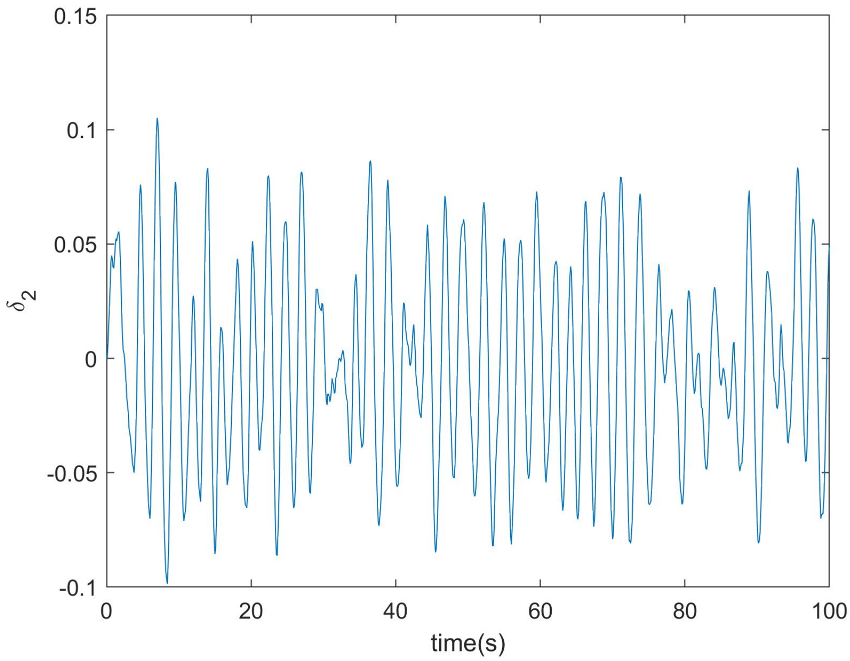

Figure 6.15 Trajectory of δ_2 on $[0, 100s]$

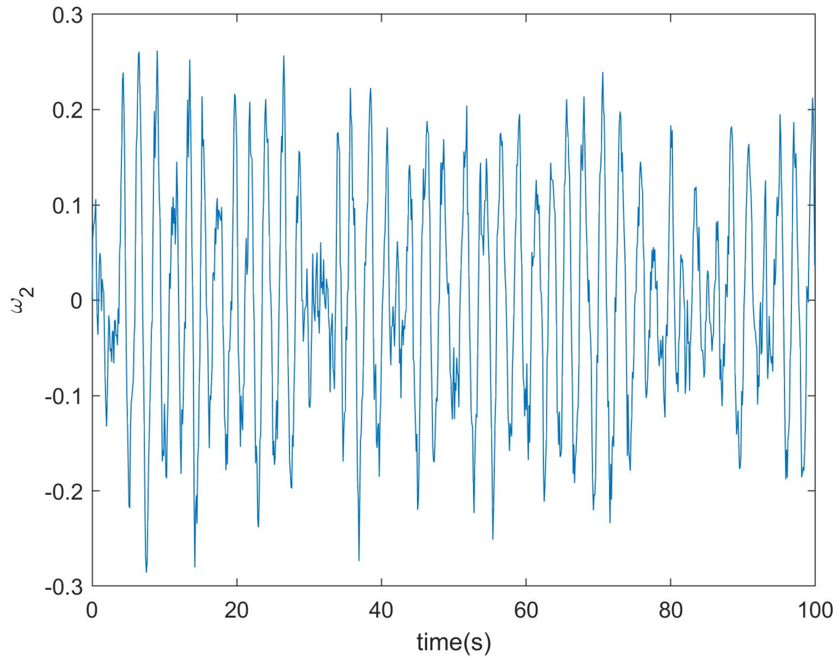

Figure 6.16 Trajectory of ω_2 on $[0, 100s]$

The system model in section (6.3) is approximated with varying Δt to present the effects of discretization on the algorithm. The trajectory plots of x are as follows:

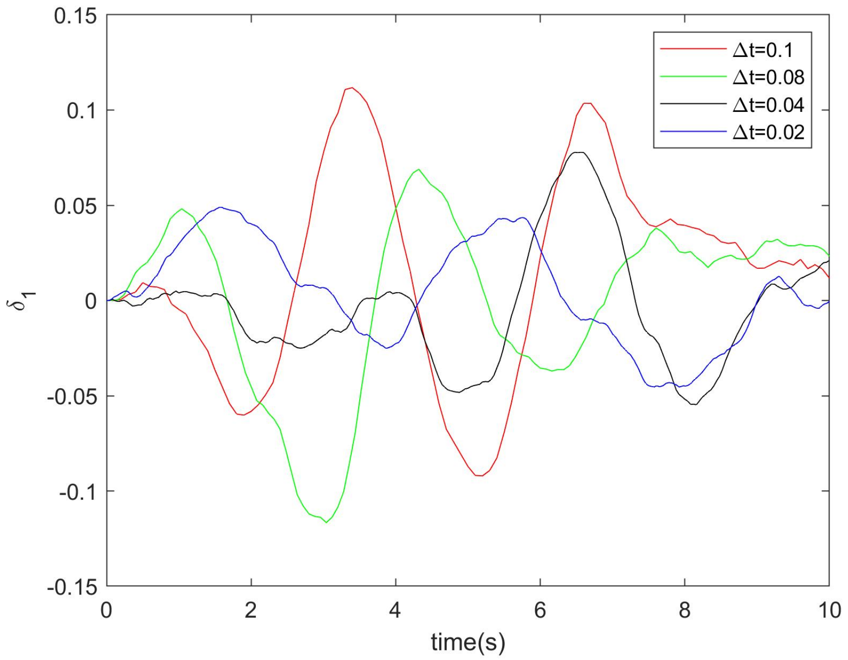

Figure 6.17 Trajectories of δ_1 for varying Δt

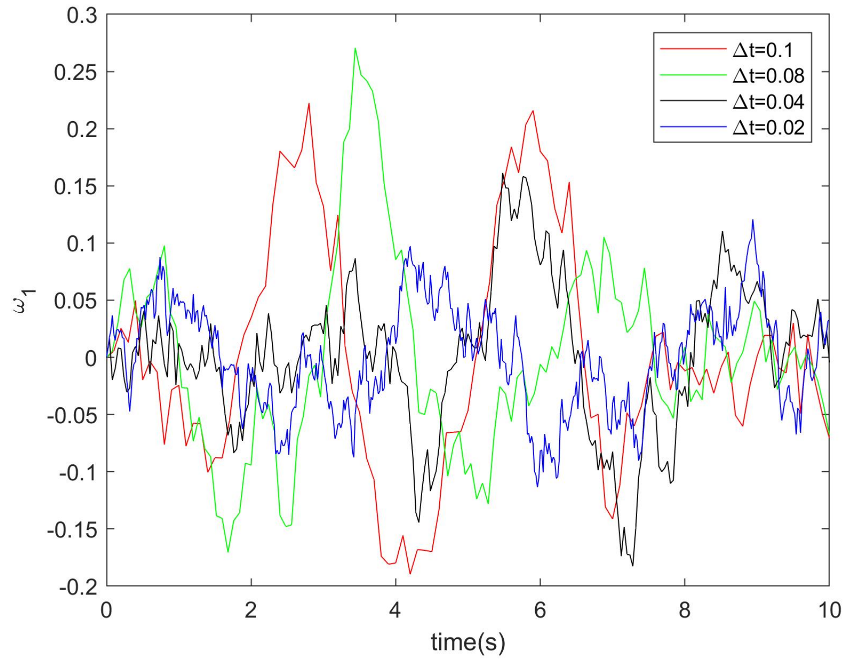

Figure 6.18 Trajectories of ω_1 for varying Δt

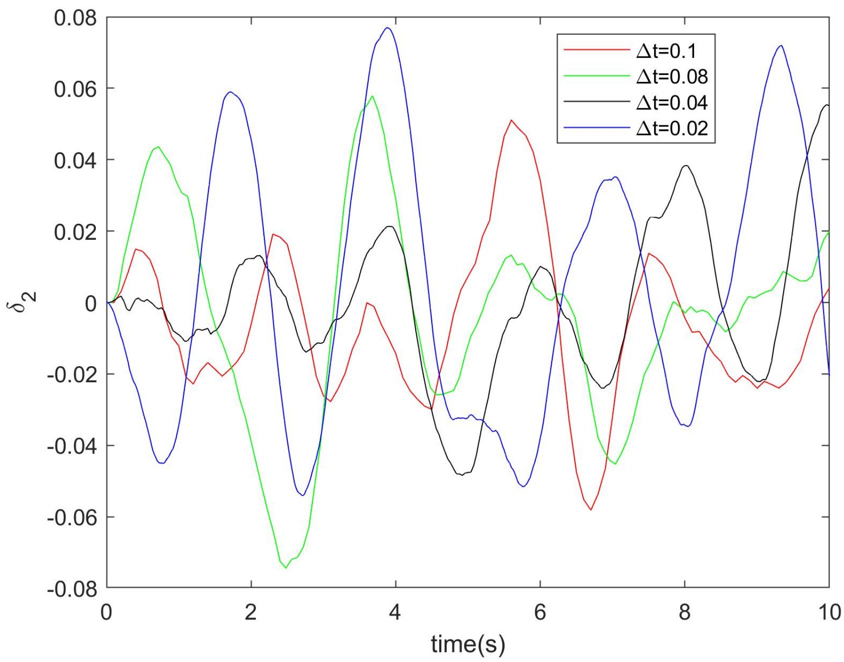

Figure 6.19 Trajectories of δ_2 for varying Δt

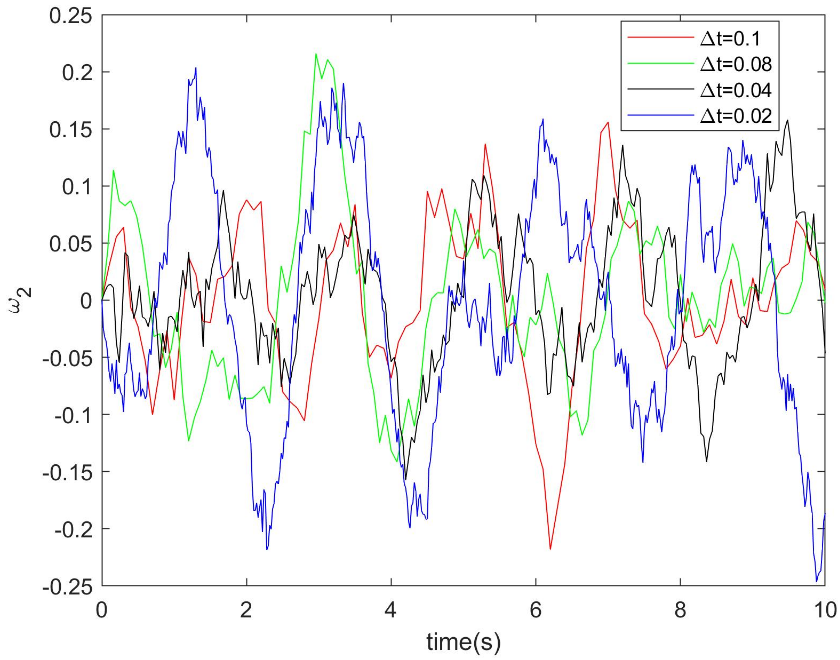

Figure 6.20 Trajectories of ω_2 for varying Δt

When the system in section (6.3) is approximated with varying the diffusion of the Brownian motion Q, the trajectory plots of x is as follows:

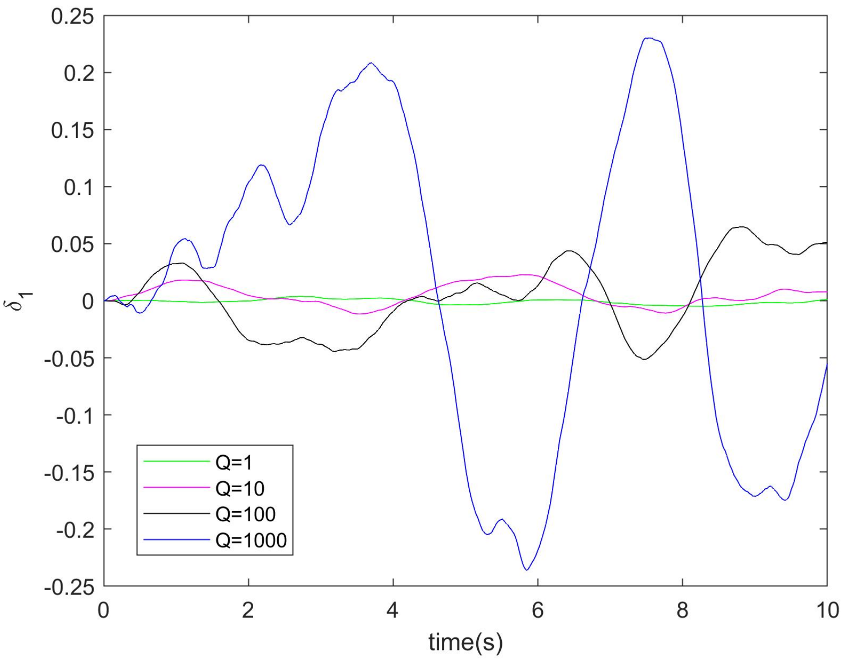

Figure 6.21 Trajectories of δ_1 for varying Q

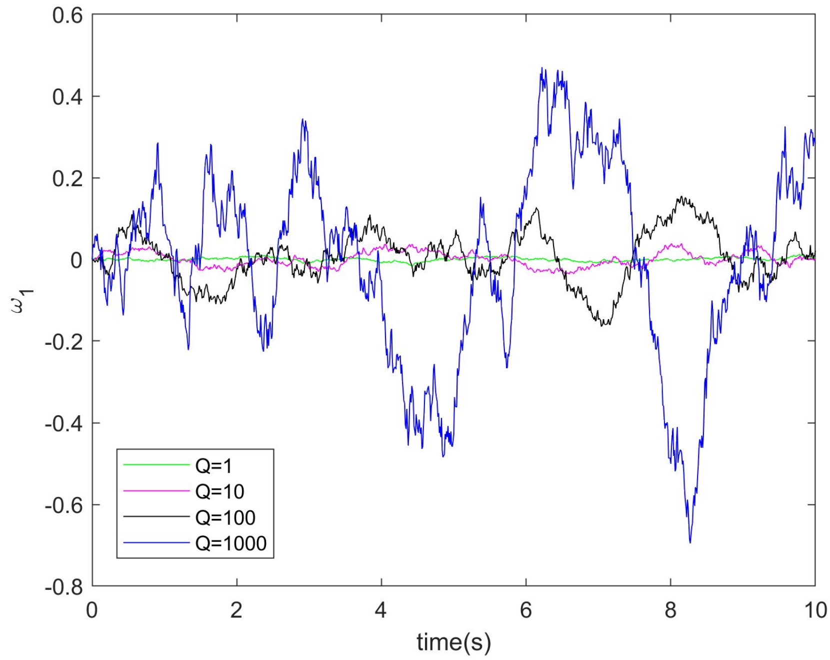

Figure 6.22 Trajectories of ω_1 for varying Q

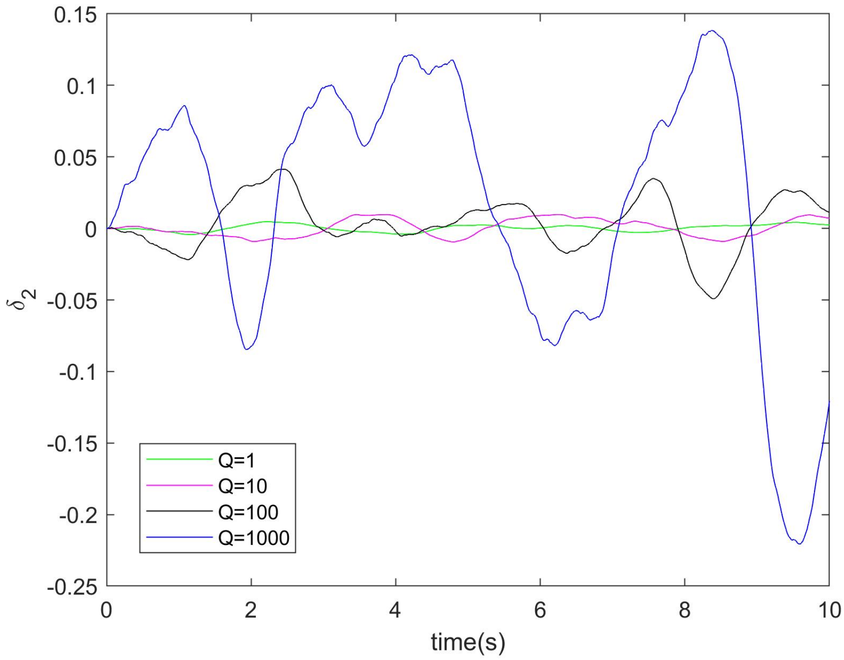

Figure 6.23 Trajectories of δ_2 for varying Q

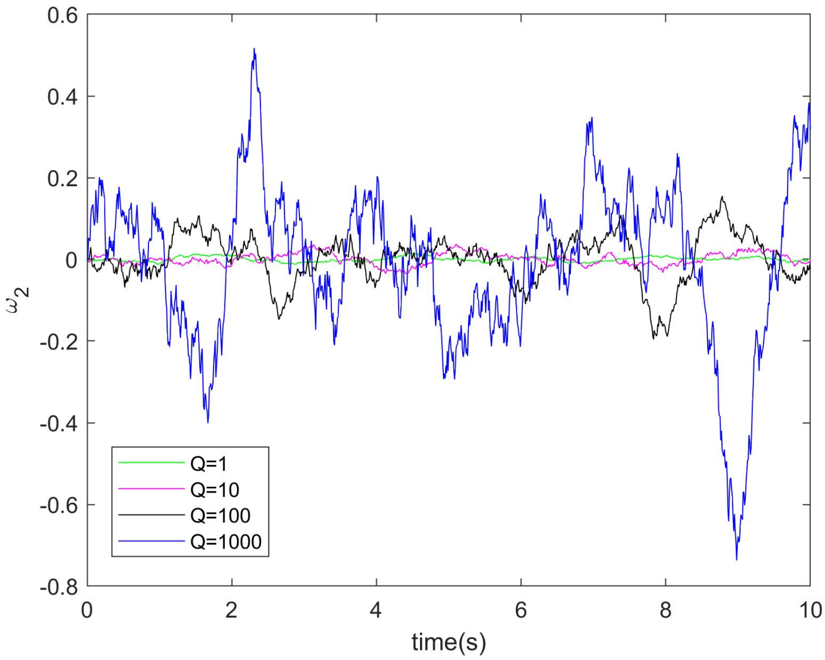

Figure 6.24 Trajectories of ω_2 for varying Q

6.6 Ito-Taylor Method

When the system model in section (6.3) is approximated using the Ito-Taylor method, the trajectory plots obtained are as follows:

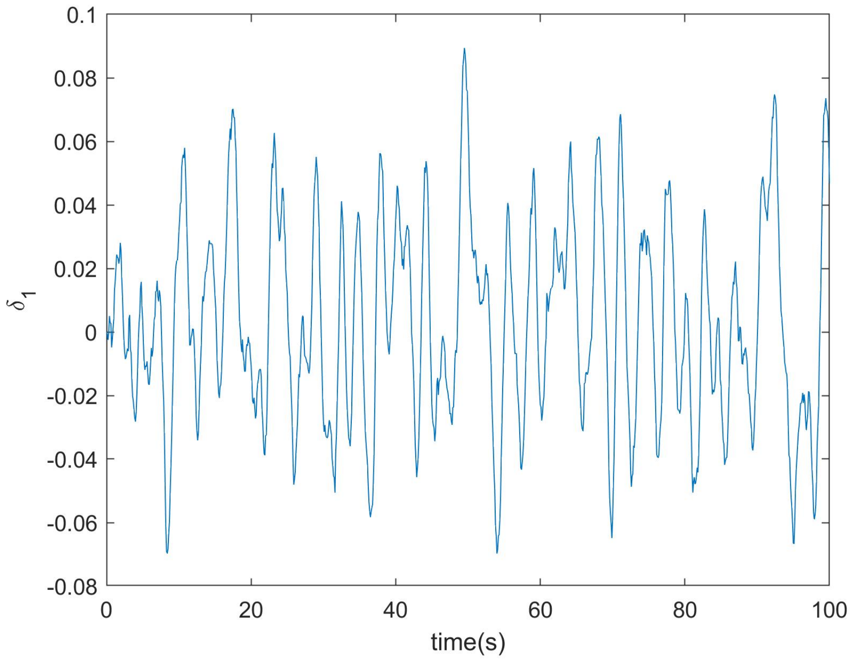

Figure 6.25 Trajectory of δ_1 on $[0, 100s]$

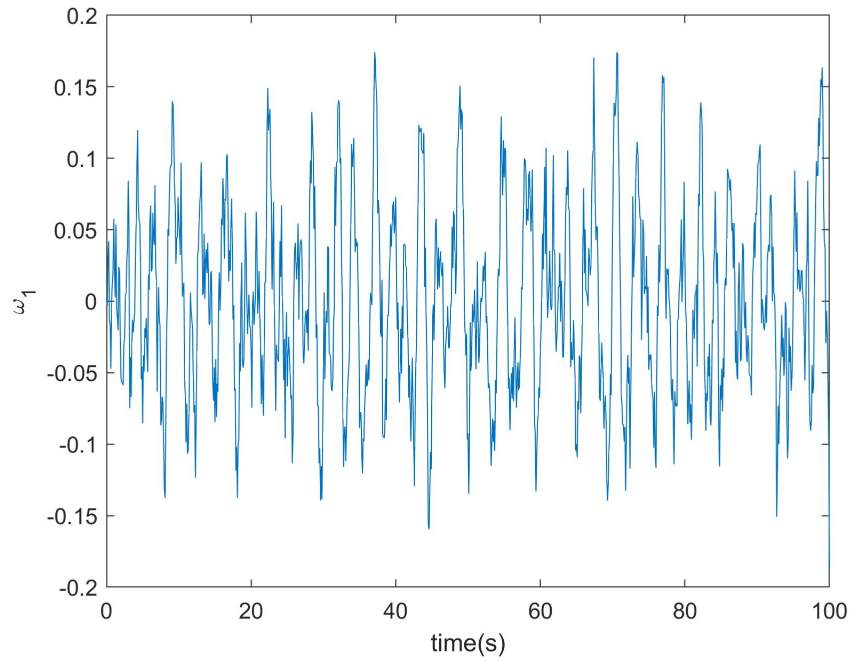

Figure 6.26 Trajectory of ω_1 on $[0, 100s]$

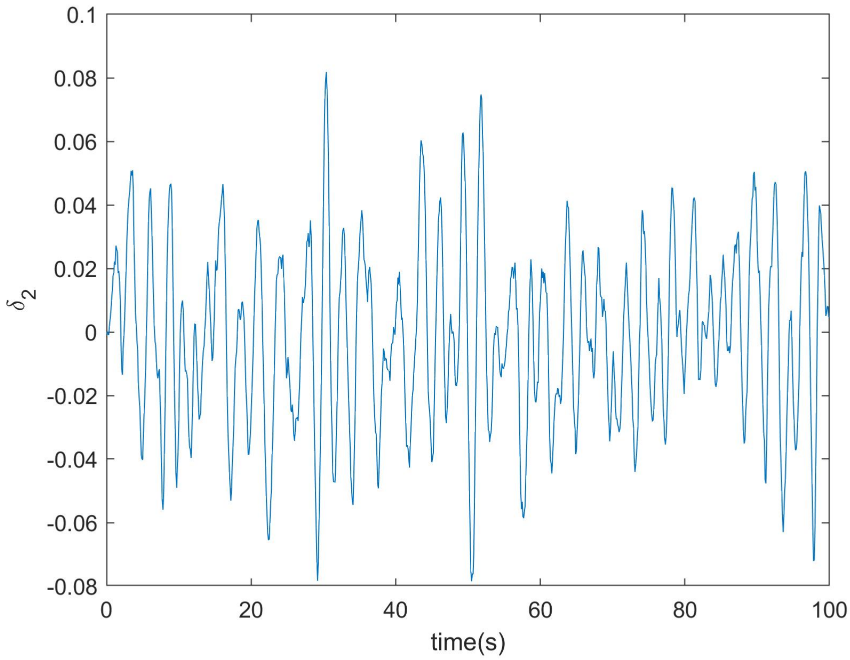

Figure 6.27 Trajectory of δ_2 on $[0, 100s]$

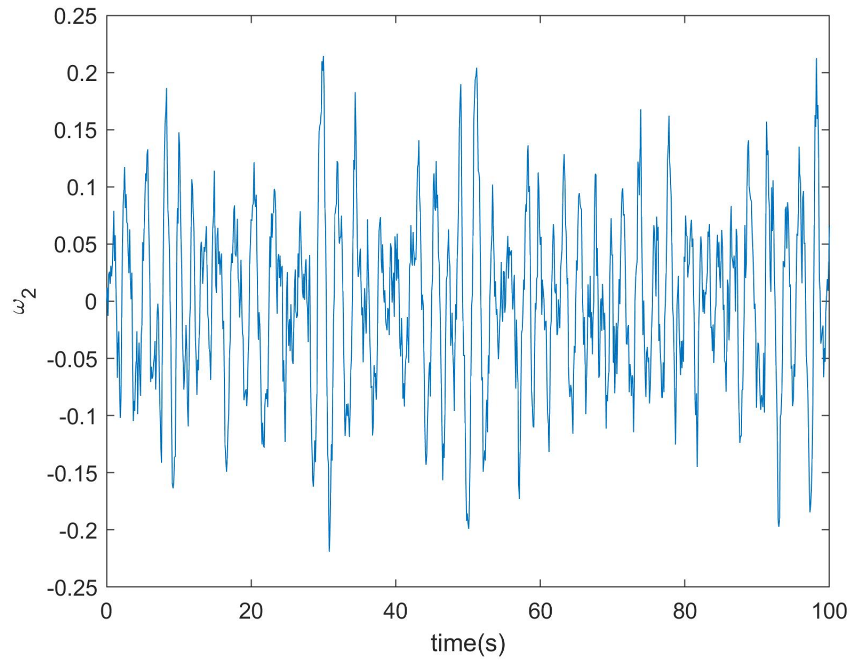

Figure 6.28 Trajectory of ω_2 on $[0, 100s]$

The system model is approximated with varying Δt to analyse the effects of discretization on the approximated algorithm. The trajectory plots of x are as follows:

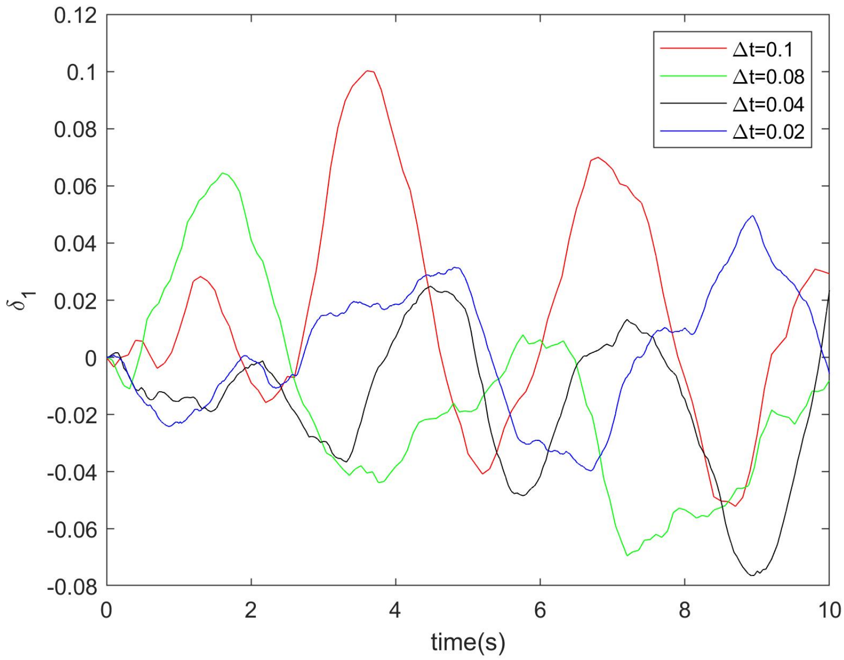

Figure 6.29 Trajectories of δ_1 for varying Δt

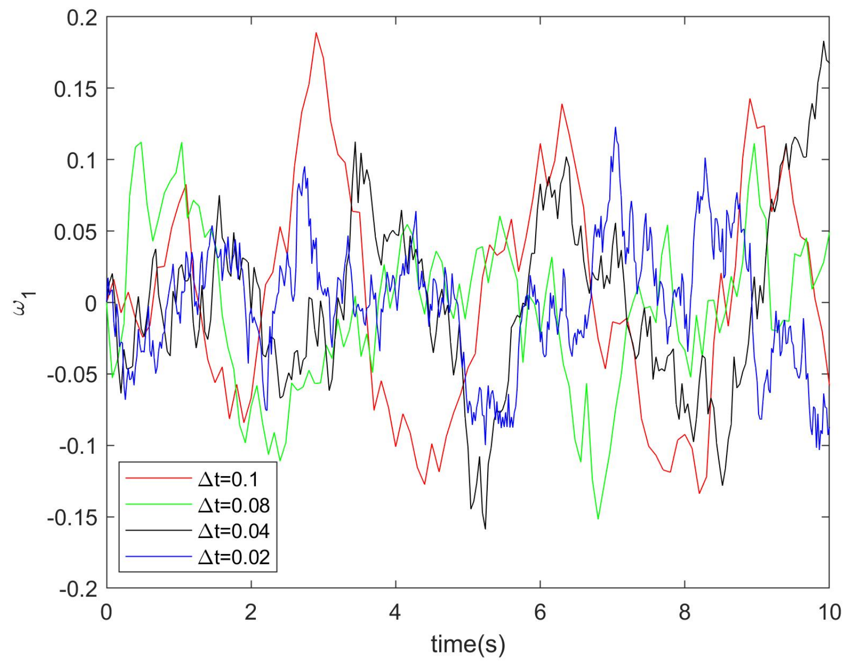

Figure 6.30 Trajectories of ω_1 for varying Δt

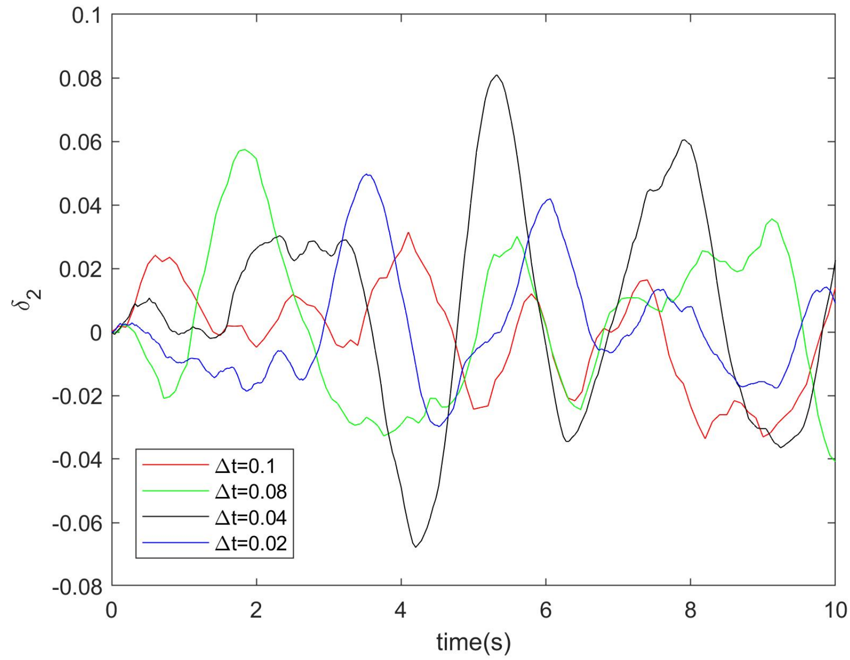

Figure 6.31 Trajectories of δ_2 for varying Δt

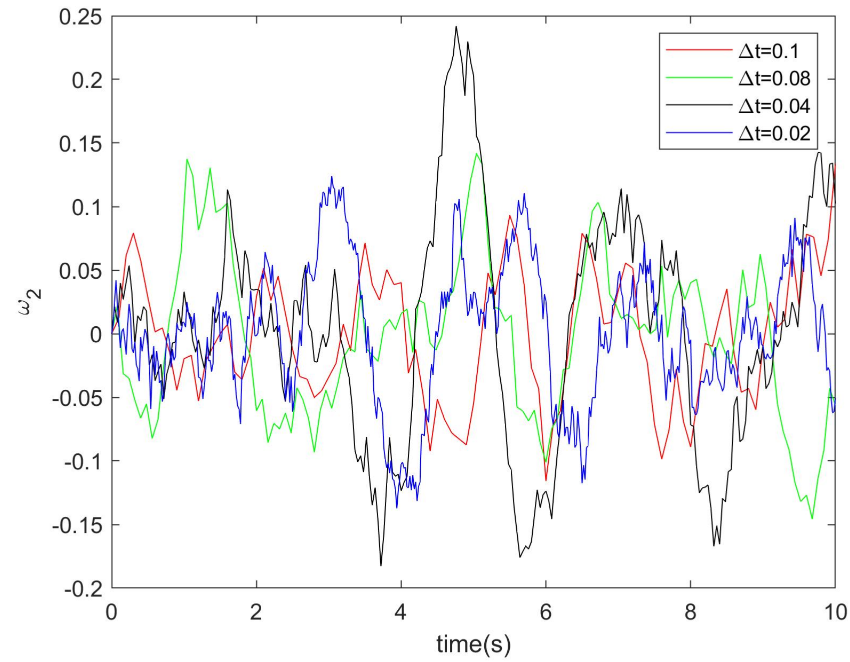

Figure 6.32 Trajectories of ω_2 for varying Δt

The system model when approximated with varying diffusion of the Brownian motion Q, gives the trajectories shown in plots below.

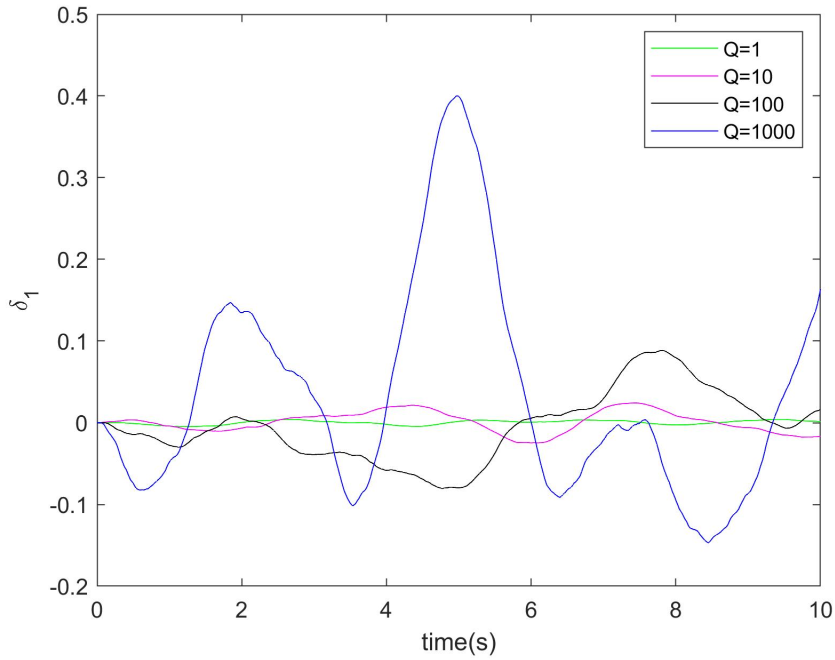

Figure 6.33 Trajectories of δ_1 for varying Q

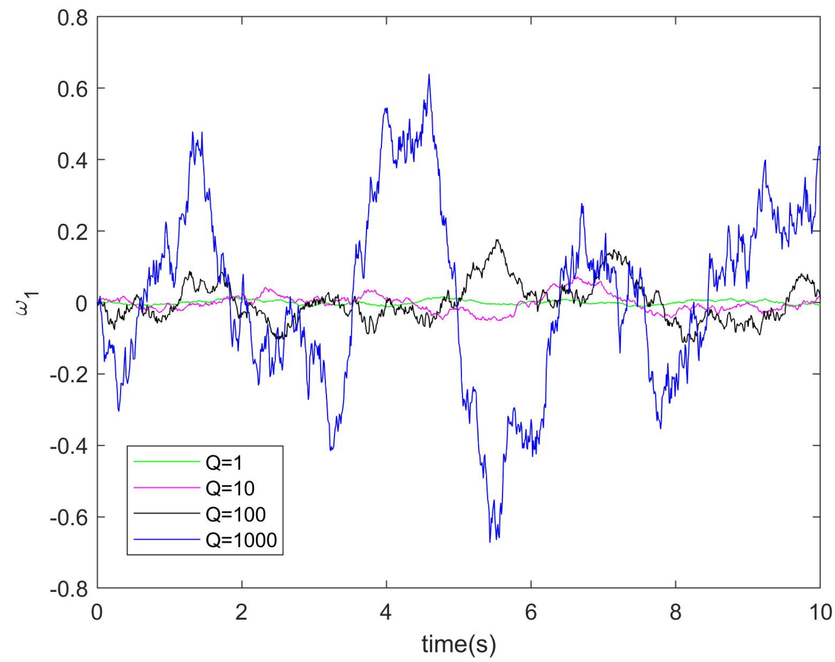

Figure 6.34 Trajectories of ω_1 for varying Q

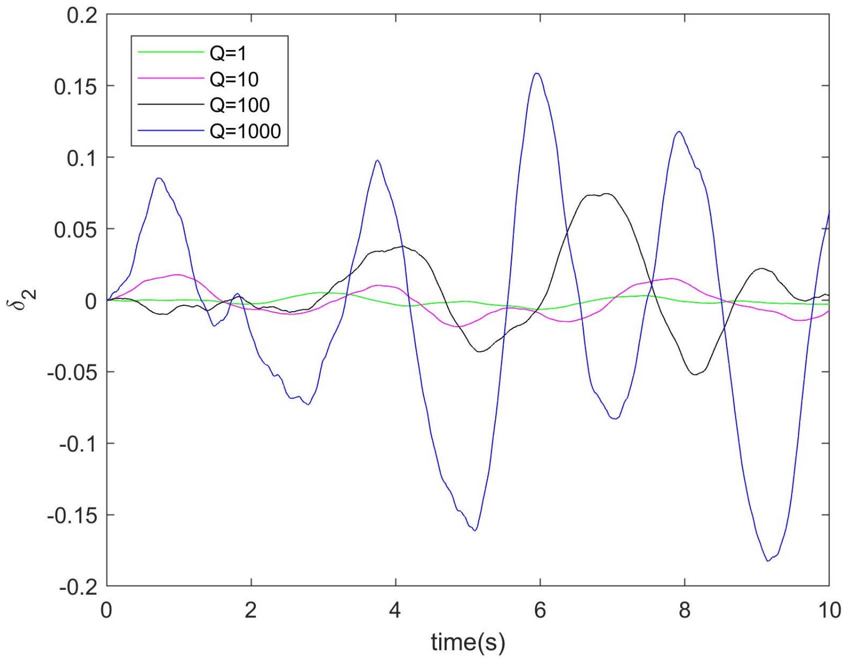

Figure 6.35 Trajectories of δ_2 for varying Q

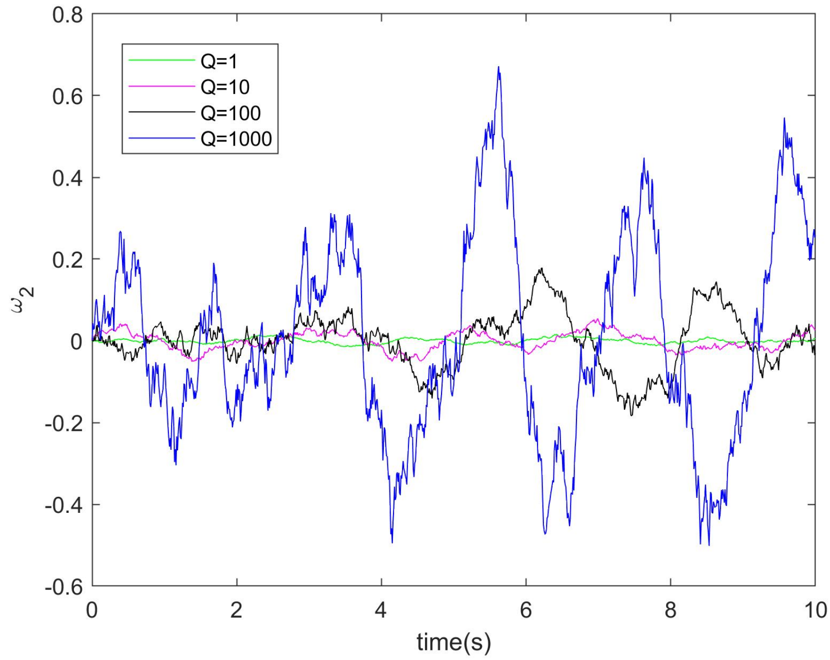

Figure 6.36 Trajectories of ω_2 for varying Q

6.7 Conclusion

The thesis is designed to give the readers a brief insight on how the power system is studied using various numerical methods that solve Stochastic Differential Equations. In our system, since the function L is independent of x and t, the partial derivatives in Milstein's method is 0. Therefore, Milstein's method is reduced to Euler-Maruyama method. The graphs are plotted with various discretization values. From these graphs we observe that when dt is too huge and we approximate x_t, the time taken is longer. In addition to that, with higher dt, the trajectories are not accurate since not all the points on the curve are included, leading to slower convergence. But when more points are included, the approximation converges faster.

In addition to these methods, there are other methods like Runge-Kutta method, stronger order of convergence Itô-Taylor methods, etc., although these methods tend to be more complex to solve for higher order systems. A more in depth study of such numerical methods can be found in Kloeden and Platen 1999 [3].

Bibliography

[1] X. Wang and K. Turitsyn. "PMU-based estimation of dynamic state Jacobian matrix". In: *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2017, pp. 1–4.

[2] Simo Särkkä and Arno Solin. *Applied Stochastic Differential Equations*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2019. DOI: 10.1017/9781108186735.

[3] P. E. Kloeden and Platen E. *Numerical Solutions to Differential Equations*. Springer, Berlin, Heidelberg, 1992. DOI: 10.1007/978-3-662-12616-5.

[4] Lawrence C. Evans. *Instructor Notes: An Introduction to Stochastic Differential Equations*. 2010.

[5] Albert Einstien. In: (1905). DOI: 10.1002/andp.19053220806.

[6] Bernt Øksendal. *Stochastic Differential Equations- An Introduction with Applications*. Springer-Verlag Berlin Heidelberg, 2003. DOI: 10.1007/978-3-642-14394-6.

[7] Simo Särkkä. *Recursive Baysian Inference on Stochastic Differential Equations*. Teknillisen korkeakoulun Laskennallisen tekniikan laboratorion julkaisuja, 2006.

[8] Martin Haugh. "Simulating Stochastic Differential Equations". In: (2017).

[9] Timothy Sauer. "Computational solution of stochastic differential equations". In: *WIREs Computational Statistics* 5.5 (2013), pp. 362–371. DOI: 10.1002/wics.1272. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.1272.

[10] T. Partal M. Bayram and G. Orucova Buyukoz. "Numerical methods for simulation of stochastic differential equations". In: *Advances in Difference Equations 2018* 17 (2018). URL: https://doi.org/10.1186/s13662-018-1466-5.

[11] Hannah Michalska. *Instructor Notes: Parameter Estimation using Lsplines in RKHS*. 2019.

[12] *Weiner Process. Gaussian White Noise.* URL: https://www.eng.tau.ac.il/~liptser/lectures/lect_new10.pdf.

[13] Desmond J. Higham. “An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations”. In: (2001).

[14] Timothy Sauer. “An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations”. In: (2001).

[15] Leslaw Socha. *Linearization Methods for Stochastic Dynamic Systems.* Springer-Verlag Berlin Heidelberg, 2008. DOI: 10.1007/978-3-540-72997-6.

[16] Henri Schurz Peter Eris Kloeden Eckhard Platen. *Numerical Solution of SDE Through Computer Experiments.* Springer-Verlag Berlin Heidelberg, 1994. DOI: 10.1007/978-3-642-57913-4.

Appendix A

Itô-Calculus and SDEs [2]

A.1 Taylor Series of ODEs

We can use Taylor series expansion as one of the methods, to find the approximate solutions of deterministic ordinary differential equations. It is quite practical in ODE numerical approximation, but is superseded by Runge-Kutta type of derivative free methods. Since the corresponding *Itô*-Taylor series solution of SDEs, in the next section, provides a useful basis for numerical methods of SDEs, and the derivation is also analogous to the ODE derivation, we will derive the Taylor series solution for ODEs first.

$$\frac{dx(t)}{dt} = f(x(t), t), x(t_0) = given, \tag{A.1}$$

which can be integrated to give,

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x(\tau).\tau) d\tau. \tag{A.2}$$

If the function f is differentiable, we can also write as $t \mapsto f(x(t), t)$ as the solution to the differential equation

$$\frac{df(x(t), t)}{dt} = \frac{\partial f}{\partial t}(x(t), t) + \sum_i f_i(x(t), t) \frac{\partial f}{\partial x_i}(x(t), t), \tag{A.3}$$

where $(x(t_0), t_0)$is the given initial condition. The integral form of this is,

$$f(x(t), t) = f(x(t_0), t_0) + \int_{t_0}^{t} \left[\frac{\partial f}{\partial t}(x(\tau), \tau) + \sum_i f_i(x(\tau), \tau) \frac{\partial f}{\partial x_i}(x(\tau), \tau) \right] d\tau. \tag{A.4}$$

The linear operator can be defined as,

$$\mathscr{L}g = \frac{\partial g}{\partial t} + \sum_i f_i \frac{\partial g}{\partial x_i} \tag{A.5}$$

and rewrite the integral equation,

$$f(x(t), t) = f(x(t_0), t_0) + \int_{t_0}^{t} Lf(x(\tau), \tau)d\tau. \tag{A.6}$$

Substituting this equation is Equation (A.2) gives,

$$\begin{aligned} x(t) = x(t_0) + \int_{t_0}^{t} [f(x(t_0), t_0) + \int_{t_0}^{t} \mathscr{L}f(x(\tau), \tau)d\tau]d\tau \\ = x(t_0) + f(x(t_0), t_0)(t - t_0) + \int_{t_0}^{t}\int_{t_0}^{t} \mathscr{L}f(x(\tau), \tau)d\tau d\tau. \end{aligned} \tag{A.7}$$

The term in the integrand on the right can be defined as a solution to the differential equation

$$\begin{aligned} \frac{d[\mathscr{L}f(x(t), t)]}{dt} &= \frac{\partial[\mathscr{L}f(x(t), t)]}{\partial t} + \sum_i f_i(x(t), t)\frac{\partial[\mathscr{L}f(x(t), t)]}{\partial x_i} \\ &= \mathscr{L}^2 f(x(t), t). \end{aligned} \tag{A.8}$$

which is in integral form,

$$\mathscr{L}f(x(t), t) = \mathscr{L}f(x_0), t_0) + \int_{t_0}^{t} \mathscr{L}^2 f(x(\tau), \tau)d\tau. \tag{A.9}$$

Substituting into the equation of $x(t)$ which gives,

$$\begin{aligned} x(t) &= x(t_0) + f(x(t_0), t)(t - t_0) \\ &+ \int_{t_0}^{t}\int_{t_0}^{t} [\mathscr{L}f(x(t_0), t_0) + \int_{t_0}^{\tau} \mathscr{L}^2 f(x(\tau), \tau)d\tau]d\tau d\tau \\ &= x(t_0) + f(x(t_0), t_0)(t - t_0) + \frac{1}{2}\mathscr{L}f(x(t_0), t_0)(t - t_0)^2 \\ &+ \int_{t_0}^{t}\int_{t_0}^{\tau}\int_{t_0}^{\tau} \mathscr{L}^2 f(x(\tau), \tau)d\tau d\tau d\tau \end{aligned} \tag{A.10}$$

Continuing this procedure ad infinitum, we obtain the following Taylor series expansion for the solution of the ODE:

$$\begin{aligned} x(t) = x(t_0) + f(x(t_0), t_0)(t–t_0) + \frac{1}{2!}\mathscr{L} f(x(t_0), t_0)(t–t_0)^2 \\ + \frac{1}{3!}\mathscr{L}^2 f(x(t_0), t_0)(t–t_0)^3 + \cdots \end{aligned} \tag{A.11}$$

From the derivation above, if we truncate the series at the nth term, the residual error is:

$$r_n(t) = \int_{t_0}^{t} \cdot \int_{t_0}^{t} \mathscr{L}^n f(x(\tau), \tau) d\tau^{n+1}, \tag{A.12}$$

which by integration by parts and mean value theorem, can be further simplified. To derive the series expansion for an arbitrary function $x(t)$, we can define it as a solution to the trivial differential equation

$$\frac{dx}{dt} = f(t), x(t_0) = given. \tag{A.13}$$

where f(t) = dx(t) /dt. Because f is independent of x, therefore,

$$\mathscr{L}^n f = \frac{d^{n+1}x(t)}{dt^{n+1}}. \tag{A.14}$$

Thus, the corresponding series becomes the classical Taylor series:

$$\begin{aligned} x(t) = x(t_0) + \frac{dx}{dt}(t_0)(t–t_0) + \frac{1}{2!}\frac{d^2x}{dt^2}(t_0)(t–t_0) \\ + \frac{1}{3!}\frac{d^3x}{dt^3}(t_0)(t–t_0)^3 + \cdot \end{aligned} \tag{A.15}$$

A stochastic differential equation can be heuristically considered as a vector differential equation of the form, which is already discussed in the previous section,

$$\frac{dx}{dt} = f(x, t) + L(x, t)w(t), \tag{A.16}$$

where $w(t)$ is a zero mean white Gaussian process. This is sometimes true, but not all the time. Therefore, in this section we shall read about what goes on with a stochastic differential equation and how we are supposed to treat them.

The problem with equation (A.16) is that it cannot be a differential equation in the traditional sense, because the ordinary theory of differential equations does not permit discontinuous functions, the white Gaussian process $w(t)$ in differential equations. The problem is not all theoretical because the solution turns out to depend on infinitesimally small differences in mathematical definitions of the noise and thus, without further restrictions the solution would not be unique even with a given realization of white noise $w(t)$. The solution to this problem is that, if we need to reduce the problem to define a new kind of integral called the Itô integral, which is an integral with respect to a stochastic process. Let us formally first integrate the differential equation from some initial time t_0 to final time t:

$$x(t) - x(t_0) = \int_{t_0}^{t} f(x(t), t)dt + \int_{t_0}^{t} L(x(t), t)w(t)dt \tag{A.17}$$

The first integral is just a normal integral with respect to time and can be defined a Riemann integral of $t \longmapsto f(x(t), t)$ or as a Lebesgue integral with respect to the Lebesgue measure, for generality.

The second integral is the one which cannot be defined as Riemann integral due to the unboundedness and discontinuity of the white noise process. In the Riemannian sense, the integral would be defined as the following kind of limit :

$$\int_{t_0}^{t} L(x(t), t)w(t)dt = \lim_{n\to\infty} \sum_{k} L(x(t_k^*), t_k^*)w(t_k^*)(t_k + 1 - t_k) \tag{A.18}$$

where $t_0 < t_1 < \cdots < t_n = t$ and $t_k^* \in [t_k, t_k + 1]$. In the context of Riemann integrals so called upper and lower sums are defined as the selections of t_k^* such that the integrand $L(s(t_k^*), t_k^*)w(t_k^*)$ has its maximum and minimum values respectively. The Riemann integral is defined if the upper and lower sums converge to the same value, which then defined to be the value of the integra. In the case of white noise, since it is not bounded and takes arbitrarily small and large values at every finite interval, the Riemann integral does not converge.

We could also attempt to define it as a Stieltjes integral which is more general than the Riemann integral. To define it, we need to interpret the increment $w(t)dt$ as increment of another process $\beta(t)$ such that the integral becomes:

$$\int_{t_0}^{t} L(x(t), t)w(t)dt = \int_{t_0}^{t} L(x(t), t)d\beta(t) \tag{A.19}$$

Turns out, that the Brownian motion discussed in the previous section is a suitable process for this purpose.

Unfortunately, the definition of the latter integral in Equation(A.18) in terms of increments of Brownian motion as in Equation (A.19) does not solve our existence problem. The problem is the discontinuous derivative of $\beta(t)$ which makes it too irregular for the defining sum of Stieltjes integral to converge. The same issue arises with Lebesgue integral. These integrals are essentially defined as limits of the form:

$$\int_{t_0}^{t} L(x(t),t)d\beta = \lim_{n\to\infty} \sum_k L(x(t_k^*),t_k^*)[\beta(t_k+1)-\beta(t_k)] \tag{A.20}$$

where $t_0 < t_1 < \cdots < t_n$ and $t_k^* \in [t_k, t_k+1]$. The main problem in both of these definitions is that they would require the limit to be independent of the position on the interval $t_k^* \in [t_k, t_k+1]$. But for integration with respect to Brownian motion this is not the case. Therefore, the two integral definitions, Stieltjes or Lebesgue, do not work either.

The solution is the Itô stochastic integral which is based on the observation that if we fix the choice to $t_k^* = t_k$, then the limit becomes unique. The Itô integral can thus be defined as the limit:

$$\int_{t_0}^{t} L(x(t),t)d\beta(t) = \lim_{n\to\infty} \sum_k L(x(t_k),t_k)[\beta(t_k+1)-\beta(t_k)] \tag{A.21}$$

which is a sensible definition of the stochastic integral required for the SDE.

The stochastic differential equation (3.21) can be defined as, corresponding to the Itô integral equation

$$x(t)-x(t_0) = \int_{t_0}^{t} f(x(t),t)dt + \int_{t_0}^{t} L(x(t),t)d\beta(t), \tag{A.22}$$

which should be true for arbitrary $t_0 andt$. We can choose the integration limits in Equation (A.22) to be $tandt + dt$, where dt is “small”, we can write the equation in the differential form as

$$dx = f(x,t)dt + L(x,t)d\beta, \tag{A.23}$$

which is the shorthand representation of the integral equation and most often used in literature on stochastic differential equation. We can now formally divide by dt to obtain a differential equation:

$$\frac{dx}{dt} = f(x,t) + L(x,t)\frac{d\beta}{dt}, \tag{A.24}$$

which shows that here white noise can be interpreted as the formal derivative of Brownian motion. However, due to non-classical transformation properties of the Itô differentials, one has to be careful in working with such formal manipulations.

It is now easy to see why we are not permitted to consider more general differential equations of the form:

$$\frac{dx(t)}{dt} = f(x(t), w(t), t). \tag{A.25}$$

where the white noise $w(t)$ enters the system through a non-linear transformation. We cannot rewrite this equation as a stochastic integral with respect to a Brownian motion and thus, cannot define the mathematical meaning of this equation. White noise generally should not be thought as an entity as such, but only exists as the formal derivative of Brownian motion. Therefore, only linear functions of white noise have a meaning whereas non-linear functions do not.

Let us now see how Itô integrals are often treated in stochastic analysis. In the above solution, we have only considered stochastic integration of the term $L(x(t), t)$, but the definition can be extended to arbitrary Itô processes $\Theta(t)$, which are "adapted" to the Brownian motion $\beta(t)$ to be integrated over. The meaning of "adapted" here means that $\beta(t)$ is the only stochastic "driving force" in $\Theta(t)$ in the same sense that $L(x(t), t)$ was generated as function of $x(t)$, which in turn is generated through the differential equation, where the only stochastic driver is the Brownian motion. This adaptation can be denoted by including the "event space element" ω as argument to the function $\Theta(t, \omega)$ and Brownian motion $\beta(t, \omega)$. The resulting Itô integral is then defined as the limit

$$\int_{t_0}^{t} \Omega(t, \omega) d\beta(t, \omega) = \lim_{n \to \infty} \sum_{k} \Omega(t_k, \omega)[\beta(t_k + 1, \omega) - \beta(t_k, \omega)]. \tag{A.26}$$

The definition is slightly more complicated, but the principle is the same as above. This kind of analysis requires us to use the full measure theoretical formulation of Itô stochastic integral which we do not do here. But in the next section we will see the Itô formula.

A.2 Itô Formula

Consider the stochastic integral

$$\int_{0}^{t} \beta(t) d\beta(t) \tag{A.27}$$

where $\beta(t)$ is a standard Brownian motion, that is, scalar Brownian motion with diffusion $Q = 1$. Based on the ordinary calculus we would expect the value of this integral to be $\beta^2(t)/2$, but that is wrong. If we select a partition $0 = t_0 < t_1 < \cdots < t_n = t$, we get by rearranging the terms

$$
\begin{aligned}
\int_0^t d\beta(t) &= \lim \sum_k \beta(t)k)[beta(t_k+1) - \beta(t_k)] \\
&= \lim \sum_k \left[-\frac{1}{2}(\beta(t_k+1) - \beta(t_k))^2 + \frac{1}{2}(\beta^2(t_k+1) - \beta^2(t_k))^2 \right] \\
&= -\frac{1}{2}t + \frac{1}{2}\beta^2(t)
\end{aligned}
\tag{A.28}
$$

where we have used the result that the limit of the first term is $\lim \sum_k (\beta(t_k+1) - \beta(t_k))^2 = t$. The Itô differential of $\beta^2(t)$ is analogously

$$
d\left[\frac{1}{2}\beta^2(t)\right] = \beta(t)d\beta(t) + \frac{1}{2}dt, \tag{A.29}
$$

not $\beta(t)d\beta(t)$ as we might expect. This is a consequence and also a drawback of the selection of the fixed $t_k^* = t_k$. The generalized rule for calculating the Itô differentials and thus Itô integrals can be summarized as the following Itô formula, which corresponds to the chain rule in ordinary calculus.

Although the Itô formula above is defined only for scalar ϕ, it is obviously works for each of the components of a vector values function separately and thus includes the vector case also. Note that every Itô process has a representation as the solution of a SDE of the form

$$
Dx = f(x,t)dt + L(x,t)d\beta \tag{A.30}
$$

and the explicit expression for the differential in terms of the function $f(x,t)$ and $L(x,t)$ could be derived by substituting the above equation for dx in the Itô formula.

The Itô formula can be conceptually derived by Taylor series expansion:

$$
\begin{aligned}
\phi(x+dx, t+dt) = \phi(x,t) &+ \frac{\partial \phi(x,t)}{\partial t}dt + \sum_i \frac{\partial \phi(x,t)}{\partial x_i}dx_i \\
&+ \frac{1}{2}\sum_{ij}\left(\frac{\partial^2 \phi}{\partial x_i \partial x_j}\right) dx_j dx_j + \dots
\end{aligned}
\tag{A.31}
$$

that is, to the first order in dt and second order in dx we have

$$
\begin{aligned}
d\phi &= \phi(x+dx, t+dt) - \phi(x,t) \\
&\approx \frac{\partial \phi(x,t)}{\partial t}dt + \sum_i \frac{\partial \phi(x,t)}{\partial x_i}dx_i + \frac{1}{2}\sum_{ij}\left(\frac{\partial^2 \phi}{\partial x_i \partial x_j}\right) dx_i dx_j.
\end{aligned}
\tag{A.32}
$$

In deterministic case we could ignore the second order and higher order terms, because $dxdx^T$ would already be of the order dt^2. Thus the deterministic counterpart is

$$d\phi = \frac{\partial \phi}{\partial t} dt + \frac{\partial \phi}{\partial x} dx. \tag{A.33}$$

But in the stochastic case we know that $dxdx^T$ is potentially of the order dt, because $d\beta d\beta^T$ is of the same order. Thus, we need to retain the second order term also.

A.3 Explicit Solutions of Linear SDEs

In this section, we derive the full solution to a general time-varying linear stochastic differential equation. The time-varying multidimensional SDE is assumed to have the form:

$$dx = F(t)xdt + u(t)dt + L(t)d\beta \tag{A.34}$$

where $x \in \mathbb{R}^n$ is a Brownian motion.

We continue by defining a transition matrix $\Psi(\tau, t)$. We multiply the above SDE with the integrating factor $\Psi(t_0, t)$ and rearranging the terms,

$$\Psi(t_0, t)dx - \Psi(t_0, t)F(t)xdt = \Psi(t_0, t)u(t)dt + \Psi(t_0, t)L(t)d\beta \tag{A.35}$$

The Itô formula gives us,

$$d[\Psi(t_0, t)x] = -\Psi(t, t_0)F(t)xdt + \Psi(t, t_0)dx \tag{A.36}$$

Thus, the SDE can be rewritten as

$$d[\Psi(t_0, t)x] = \Psi(t_0, t)u(t)dt + \Psi(t_0, t)L(t)d\beta \tag{A.37}$$

here the differential is a Itô differential. Now we apply Itô integration from $t_0 tot$ which gives,

$$\Psi(t_0, t)x(t) - \Psi(t_0, t_0)x(t_0) = \int_{t_0}^{t} \Psi(t_0, \tau)u(\tau)d\tau + \int_{t_0}^{t} \Psi(t_0, \tau)L(\tau)d\beta(\tau), \tag{A.38}$$

which can further be written in the form

$$x(t) = \Psi(t, t_0)x(t_0) + \int_{t_0}^{t} u(\tau)d\tau + \int_{t_0}^{t} \Psi(t, \tau)L(\tau)d\beta(\tau), \tag{A.39}$$

which is the desired solution.

In the case of LTI SDE,

$$dx = Fxdt + Ld\beta \tag{A.40}$$

where F and L are constant and β has a constant diffusion Q, the solution becomes,

$$x(t) = exp(F(t - t_0))x(t_0) + \int_{t_0}^{t} exp(F(t - \tau))Ld\beta(\tau), \tag{A.41}$$

This is the solution what we would expect if we formally replaced $w(\tau)d\tau$ with $d\beta(\tau)$ in the deterministic solution. It is just the usage of Itô formula which resulted as we expect the deterministic differentiation would, however, we cannot expect to get the right solution in the non-linear case with this kind of replacement.