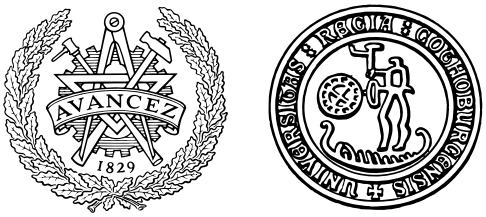
Thesis for the Degree of Licentiate of Technology

# Recent Advances in the Numerical Solution of Stochastic Differential Equations

Anders Muszta

CHALMERS | GÖTEBORG UNIVERSITY




Department of Mathematical Statistics
Chalmers University of Technology
and Göteborg University
SE-412 96 Göteborg, Sweden

Göteborg, Sweden 2003

Recent Advances in the Numerical Solution of Stochastic Differential Equations

Anders Muszta



ISSN 0347-2809
Preprint number 2003:2
Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Chalmers University of Technology
Göteborg, Sweden 2003

Abstract

Broadly speaking, stochastic differential equations (SDE) are used to model the evolution of systems influenced by random effects. The *strong solution* of the SDE allows for the detailed study of evolutionary scenarios, and the *weak solution* the study of distributional properties.

This paper is a survey devoted to numerical methods that are used to obtain approximate solutions, both strong and weak, of stochastic differential equations.

Keywords: Adaptive method; Lie group method; Stochastic differential equation; Strong approximation; Symplectic method; Waveform relaxation method; Weak approximation.

# Acknowledgements

I would like to thank the following persons:

My supervisor, Professor Patrik Albin, who has always been available to answer my questions, and to check drafts of this thesis. For this, and much more, I am very grateful.

Professors Bo Berndtson, Ivar Gustafsson and Lyudmila Turowska at Matematiska Vetenskaper, and Professor Roger Petterson at Växjö University, for helpful suggestions on different topics addressed in this thesis.

My colleagues at the department of Mathematical Statistics, for creating a friendly work environment.

My family, for their unfailing love and support.

Anders Muszta, February 2003

# Contents

# Chapter 1

# Introduction

The main objective of this thesis is to provide the reader with the underlying ideas of some of the state of the art techniques in numerical solutions of stochastic differential equations. Some of the techniques presented here are so new that only a few articles have been written on them.

Only techniques that have been presented in the scientific journals from 1992 and onward are presented here. The reason for choosing 1992 as a year of "departure", is that this is the year of the publication of the now standard work on numerical methods for SDE, by Kloeden and Platen [29].

Upon reading the material presented here, it will be noted that in many cases the novel techniques are not so "novel" after all. As is often the case, what is new in the field of numerical solution of stochastic differential equations has been employed for some time in the numerical solution of ordinary differential equations. Many such techniques have proven themselves to be adaptable to stochastic differential equations, if not directly then through a suitable modification.

# Chapter 2

# Ordinary and stochastic integration

## 2.1 Modelling changes

Upon observing the world around us, one thing stands out: Everything changes.

Mathematics tries to describe the world and naturally, in order to do this, it has to describe these changes. The mathematical concept that describes changes is that of a *derivative*. These changes usually has some cause, and the mathematical model that tries to capture this idea of cause and effect is that of a differential equation. If the causes have a random nature, then the mathematical model will be a *stochastic differential equation*.

In order to study the effect of a specific cause, the stochastic differential equation has to be *solved*. The mathematical tool for doing this is the *stochastic integral*, just as the ordinary (Riemann-Stieltjes) integral is used to solve an ordinary differential equation. Sometimes, but very rarely, the solution of a stochastic differential equation can be expressed in closed form. In most cases however, the best one can hope for is to find an approximate solution, using a numerical method.

This section illustrates the use of ordinary and stochastic differential equations by means of an example with a financial flavour: The evolution of a bank account.

### 2.1.1 Ordinary differential equations

Suppose that you open a bank account into which, at the time $t = 0$, you deposit the amount $S_0$. Then you just let the money stay in the account and

accumulate interest at a constant interest rate, $r$. Let $S(t)$ be the balance of your account at time $t$. Then the evolution of $S(t)$ can be modelled by the ordinary differential equation:

$$\begin{aligned} dS(t) &= rS(t)\,dt \\ S(0) &= S_0 \end{aligned}$$

The evolution of such an account looks like in Figure 2.1.

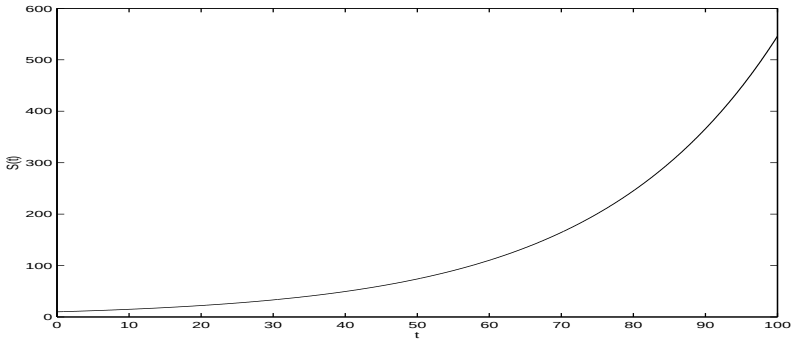



Figure 2.1: Evolution of deterministic account

Call an account of the kind described above a deterministic account. Using this model, you may want to answer questions like:

- How long do you have to wait until a certain amount has accumulated on the account?
- How much money will you have at a specified time?

### 2.1.2 Stochastic differential equations

Now consider a similar situation, where at time $t = 0$ you open the account by depositing the amount $S_0$, and the bank gives you a fixed interest rate, $r$. When using the account, you are only allowed to deposit or withdraw *small* amounts of money at any time. A model for the evolution of such an account might be:

$$\begin{aligned} dS(t) &= rS(t)\,dt + r\,dW_t \\ S(0) &= S_0 \end{aligned} \tag{2.1}$$

Here, $W_t = W(t),\ t \geq 0$, is the so called *Wiener process*, also known as *Brownian motion*. In this example, $W$ accounts for the small deposits and the small withdrawals allowed. For a definition of the Wiener process, see

page 23.

The equation (2.1) is a *stochastic differential equation* (SDE), and it is interpreted in the following integrated form:

$$S(t) = S_0 + \int_0^t rS(s)\,ds + \int_0^t r\,dW_s. \tag{2.2}$$

The second integral in (2.2) is a so called *stochastic integral*. It will be discussed on page 11.

If more than one person has access to the account, more Wiener processes are used. If for example *two* people have access, then the equation will be

$$\begin{aligned} dS(t) &= rS(t)\,dt + r\,dW_1(t) + r\,dW_2(t) \\ S(0) &= S_0 \end{aligned}$$

A typical evolution of a Wiener account with only one holder, may look like in Figure 2.2.

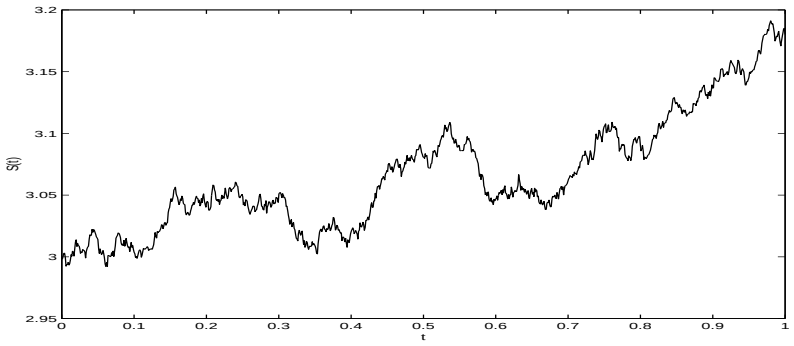



Figure 2.2: Evolution of Wiener account

If you want to model an account with very big fluctuations, then the Wiener process is not adequate any more. Instead you may want to consider using *Lévy processes*. These processes allow jumps to occur. The stochastic differential equation modelling such a situation is the same as that used for the model allowing only small changes, except that the process $W$ has been replaced with a Lévy process $L_t = L(t),\ t \geq 0$:

$$\begin{aligned} dS(t) &= rS(t)\,dt + r\,dL_t \\ S(0) &= S_0 \end{aligned}$$

Such an account is referred to as a Lévy account.

The class of Lévy processes, i.e. processes with stationary and independent increments, is very large. It includes Wiener processes, as well as processes that allow jumps and more "extreme behaviour" to occur, which does not the Wiener processes.

One realisation of a Lévy account is shown in Figure 2.3.

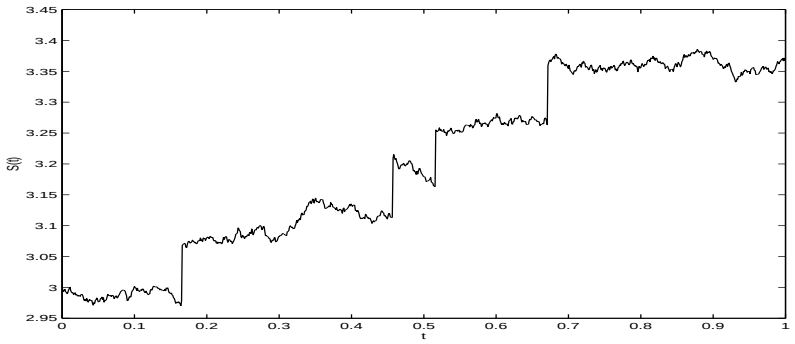



Figure 2.3: Evolution of Lévy account

Using the Wiener account, or the Lévy account, you may want to answer the same questions as those posed for the deterministic account. The stochastic accounts have an element of chance in them. Consequently it is no longer possible to tell with a single graph *exactly* how much money there is in the account at a certain time in advance. Nor is it possible to say for precisely how long you have to wait in order for the balance of the account to reach a specified level, as these characteristics are stochastic. What you on the other hand *can* compute is, for example, the *expected* balance of the account, and the *expected* time until a certain amount is reached.

The way to study a Wiener account goes via integrating the differential equations modelling the evolution of the account. The deterministic account is modelled by an ordinary differential equation, so the answers concerning the deterministic account are obtained through deterministic integration (Riemann integration). The Wiener account is modelled by a stochastic differential equation, and the answers concerning the Wiener account are obtained through *stochastic integration* (Itô integration).

## 2.2 Deterministic integration

In what follows we highlight some results from the theory of Riemann-Stieltjes integration, that are important in the theory of stochastic integration.

### 2.2.1 Riemann-Stieltjes integration

See for example Kannan [27] for the facts presented here on the Riemann-Stieltjes integral.

**Definition 2.2.1** *A function* $f : [0,1] \to \mathbb{R}$ *has* bounded $p$-variation, *for some* $p > 0$, *if*

$$\sup \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^p < \infty.$$

*Here the supremum is taken over all partitions* $0 = t_0 < t_1 < \ldots < t_n = 1$ *of* $[0,1]$.

**Theorem 2.2.1** *Let* $f, g : [0,1] \to \mathbb{R}$ *be two functions. The Riemann-Stieltjes integral* $\int_0^1 f(t)\, dg(t)$ *exists if the following conditions hold:*

1. *$f$ and $g$ do not have discontinuities at the same points in* $[0,1]$;
2. *$f$ has bounded $p$-variation and $g$ has bounded $q$-variation, for some $p > 0$ and $q > 0$ satisfying*

$$\frac{1}{p} + \frac{1}{q} > 1.$$

**Theorem 2.2.2** *If, for any* $f \in C([0,1])$, $\int_0^1 f(t) dg(t)$ *exists as a Riemann-Stieltjes integral, then $g$ is of bounded variation (i.e. of bounded 1-variation).*

### 2.2.2 Interpretation of the concept of bounded variation

The concept of bounded variation is closely related to the concept of *rectifiability*. Rectifiability is an important concept in geometric measure theory. See for example the books by Mattila [38] and Falconer [10] on this subject.

**Definition 2.2.2** *Let* $f : [a,b] \to \mathbb{R}$ *be a function and* $\{\pi_n\}_{n \in \mathbb{N}}$ *a sequence of partitions of* $[a,b]$,

$$\pi_n : a = t_0^{(n)} < t_1^{(n)} < \ldots < t_{N_n}^{(n)} = b.$$

*For each $n$, the points* $\{f(t_k^{(n)})\}_{k=1}^{N_n}$ *determine an inscribed polygon to the graph of $f$. The length of this polygon is*

$$\sum_{k=1}^{N_n} |f(t_k^{(n)}) - f(t_{k-1}^{(n)})|.$$

*The length of the curve of $f$ is defined by*

$$\sup_{\pi_n} \sum_{k=1}^{N_n} |f(t_k^{(n)}) - f(t_{k-1}^{(n)})|.$$

*If this length is finite, then the curve defined by $f$ is said to be* rectifiable.

*Upon identifying the function $f : [a, b] \to \mathbb{R}$ with its curve, if the curve defined by $f$ is rectifiable, one says that $f$ is rectifiable.*

When comparing the above definition with that of a function of bounded variation, it becomes clear that *a function is rectifiable if and only if it is of bounded variation.* This shows that many, if not all, elementary functions, like polynomials and trigonometric functions, are of bounded variation on closed intervals in $\mathbb{R}$.

Let $W$ be the Wiener process. Then $W$ has bounded $p$-variation, for $p > 2$, on any fixed finite interval $[0, T]$. Hence the integral

$$\int_0^T f(t)\, dW_t$$

exists as a Riemann-Stieltjes integral, if the function $f$ is of bounded $q$-variation for some $q < 2$. For example, the integral

$$\int_0^T \mathrm{e}^t\, dW_t$$

exists as a Riemann-Stieltjes integral. The existence of this integral in the Riemann-Steltjes sense does not, however, indicate how the integral is to be computed.

The Wiener process does not have bounded variation, and hence its curve has infinite length (an indication of the fractal nature of the Wiener process). Equivalently, it is not rectifiable. Thus there exists a continuous function $f$ on $[0, 1]$ such that the integral $\int_0^1 f(t)\, dW_t$ does not exist as a Riemann-Stieltjes integral.

## 2.3 Stochastic integration

The concept of a stochastic differential equation was introduced for the first time in 1934 by Bernstein (1880-1968) in his paper [3].

Stochastic integration is a theory developed in order to make sense of integrals like

$$\int_0^t f(s)\, dg(s),$$

where the function $g$ is stochastic and of unbounded variation. The integral can be defined as a Riemann-Stieltjes integral for $f$ with bounded variation. The challenge for a theory of stochastic integration therefore lies in a meaningful construction when $f$ does not have bounded variation.

### 2.3.1 Construction of stochastic integrals

The construction is based on approximating the integral by a sum, where the sum is interpreted as an integral of "simple functions". These are functions that are piecewise constant over the domain of integration. Thus a simple function, $f$, typically looks something like

$$\forall t \in [0,T): \quad f(t) = \sum_{k=0}^{n} c_k I_{[t_k, t_{k+1})}(t),$$

where $0 = t_0 < t_1 < \cdots < t_{n-1} < t_{n+1} = T$ is a partition of the interval $[0,T]$ and $\{c_k\}_{k=0}^{n}$ is a sequence of constants.

The integral $\int_0^T f(t)\,dg(t)$ is defined by:

$$\int_0^T f(t)\,dg(t) = \sum_{k=0}^{n} f(\tau_k)\big(g(t_{k+1}) - g(t_k)\big),$$

where $\tau_k \in [t_k, t_{k+1})$.

So far things look exactly like the procedure of constructing the Riemann-Stieltjes integral. But here is where the resemblance ceases to exist. The Riemann-Stieltjes integral is obtained as the *unique* limit of the sum above as $\delta_n \equiv \max_{0\le k\le n}(t_{k+1} - t_k) \to 0$ as $n \to \infty$, *irrespective* of how the point $\tau_k$ is chosen in each subinterval $[t_k, t_{k+1})$ of the partition. For a stochastic integral, the limit depends on how the point $\tau_k$ is chosen.

#### The Itô integral

If, in each of the subintervals, the choice

$$\tau_k = t_k$$

is made, then the integral of a simple function $f$ becomes

$$\int_0^T f(t)\,dg(t) = \sum_{k=0}^{n} f(t_k)\big(g(t_{k+1}) - g(t_k)\big). \tag{2.3}$$

If the sum in (2.3) converges as $n \to \infty$, for a not necessarily simple function $f$, then the limit is called the *Itô integral* $\int_0^T f(t)\,dg(t)$ of $f$, with respect to $g$.

### The Stratonovich integral

If, in each of the subintervals, the choice

$$\tau_k = \frac{t_k + t_{k+1}}{2}$$

is made, then the integral of a simple function $f$ becomes

$$\int_0^T f(t) \circ dg(t) = \sum_{k=0}^{n} f\Big(\frac{t_k + t_{k+1}}{2}\Big)\big(g(t_{k+1}) - g(t_k)\big). \tag{2.4}$$

If the sum in (2.4) converges as $n \to \infty$, for a not necessarily simple function $f$, then the limit is called the *Stratonovich integral* $\int_0^T f(t) \circ dg(t)$ of $f$, with respect to $g$.

### Other constructions

The Itô and Stratonovich integrals are the most used constructions of stochastic integrals. The Stratonovich integral arises when one is approximating the integrand function $g$ with a piecewise linear function (see the Wong-Zakai theorem below). This makes it the common choice, for example in physics. The Itô integral does not depend on information about the future by construction, so this makes it the common choice in finance. These are just two of an infinitude of possible constructions of a stochastic integral.

In Gard [12], the following construction is mentioned: Let $\lambda \in [0, 1]$ be a fixed parameter. Define the stochastic integral

$$\int_0^T f(t)\, dg(t)$$

for a simple function $f$ in the following way:

$$\int_0^T f(t)\, dg(t) = \sum_{k=0}^{n} \Big(\lambda f(t_{k+1}) + (1-\lambda) f(t_k)\Big)\big(g(t_{k+1}) - g(t_k)\big). \tag{2.5}$$

If the sum in (2.5) converges as $n \to \infty$, for a not necessarily simple function $f$, then the limit is the integral $\int_0^T f(t)\, dg(t)$ of $f$, with respect to $g$. It will depend on $\lambda$. If $\lambda = 0$, it is the Itô integral, and if $\lambda = \frac{1}{2}$, it is the Stratonovich integral. For other choices of $\lambda$ than these two, other stochastic integrals will be obtained.

Thus it is evident that there are infinitely many possible constructions of stochastic integrals.

**Example 2.3.1** The stochastic integral $\int_0^t W_s\, dW_s$ has the following values according to the different constructions:

| | |
|---|---|
| Itô | $\frac{1}{2}W_t^2 - \frac{1}{2}t$ |
| Stratonovich | $\frac{1}{2}W_t^2$ |
| (2.5) | $\frac{1}{2}W_t^2 + (\lambda - \frac{1}{2})t$ |

Since the integral $\int_0^t W_s dW_s$ depends on the choice of the approximating sum, it does not exist as a Riemann-Stieltjes integral.

### 2.3.2 The Itô formula

The different constructions of stochastic integrals would not be very practical, were there not to be something corresponding to the fundamental theorem of calculus. This corresponding result is called the Itô formula, and in its simplest form it looks like this:

**Theorem 2.3.1** (ITÔ FORMULA: SIMPLE CASE) *Let $f : \mathbb{R} \to \mathbb{R}$ be a function with continuous second derivative. Then $f(W_t)$ is given by*

$$f(W_t) = f(W_0) + \int_0^t f'(W_s)\, dW_s + \frac{1}{2}\int_0^t f''(W_s)\, ds. \tag{2.6}$$

The first integral in the above theorem is a *local martingale*, with respect to the filtration

$$\big\{\sigma(W_s : s \in [0,t])\big\}_{t\geq 0}$$

induced by the Wiener process, and so it has constant expectation. (See the book [28] by Karatzas and Shreve for a definition and properties of local martingales.) On account of this, it is possible to interpret the Itô formula as expressing $f(W_t)$ as a superposition of a signal part

$$f(W_0) + \frac{1}{2}\int_0^t f''(W_s)\, ds,$$

and noise part

$$\int_0^t f'(W_s)\, dW_s.$$

### 2.3.3 A theorem of Wong and Zakai

This section describes that when interpolating the Wiener process linearly, in the limit of infinitely fine time-partitions, the stochastic differential equation driven by the interpolated process, converges to the Stratonovich solution of a corresponding process. This result was obtained by Wong and Zakai [62], [63] and refined by Ikeda and Watanabe [22].

The linearly interpolated process that approximates the Wiener process seems harmless enough, but the result by Wong and Zakai gives a startling conclusion. This is so because the Stratonovich construction of stochastic integrals is *anticipating*. This means that, in a sense, in order to model the present, the future has to be known. This is in contrast to the Itô construction which is non-anticipating, i.e. in order to model the present, all that is needed is information on the past. The linear interpolation seems to be a natural construction, but nevertheless Wong and Zakai's work forces us to conclude this seemingly unnatural dependence on the future.

**Theorem 2.3.2** *For every $T > 0$, let*

$$0 = t_1 < t_2 < \cdots < t_n < t_{n+1} = T$$

*be a partition of $[0, T]$ such that*

$$\lim_{n\to\infty} \max_{1\leq k\leq n} (t_{k+1} - t_k) = 0.$$

*Let $W : \Omega \times [0, \infty] \to \mathbb{R}^N$ be a vector of $N$ independent Wiener processes. For all $n \in \mathbb{N}$ define a sequence of processes $Z_n : \Omega \times [0, \infty] \to \mathbb{R}^N$ by:*

$$\begin{aligned} &\forall k \in \{1, \ldots, n\},\ \forall t \in [t_k, t_{k+1}] : \\ &Z_n(t) = W_{t_k} + \frac{t - t_k}{t_{k+1} - t_k}\big(W_{t_{k+1}} - W_{t_k}\big) \end{aligned}$$

*Let the sequence of processes $X_n : \Omega \times [0, \infty] \to \mathbb{R}^m$ satisfy the stochastic differential equation*

$$\begin{aligned} dX_n(t) &= b(X_n(t))\, dt + a(X_n(t))\, dZ_n(t) \\ X_n(0) &= x_0 \end{aligned}$$

*Further, let the process $X : \Omega \times [0, \infty] \to \mathbb{R}^m$ satisfy the* Stratonovich *stochastic differential equation*

$$\begin{aligned} dX(t) &= b(X(t))\, dt + a(X(t)) \circ dW_t \\ X(0) &= x_0 \end{aligned}$$

*where the functions* $a \in C_B^2(\mathbb{R}^m, \mathbb{R}_{m|n})$ *and* $b \in C^1(\mathbb{R}^m, \mathbb{R}^m)$. *Then the following holds:*

$$\forall T > 0 : \quad \lim_{n\to\infty} \mathbb{E}\Big[ \max_{t\in[0,T]} \big|X_n(t) - X(t)\big|^2 \Big] = 0.$$

Observe that the notation $\ldots \circ dW_t$ is used to denote an SDE, where the solution should be interpreted in the Stratonovich sense.

# Chapter 3

# Applications

The purpose of this chapter is to show some examples that demonstrate the diverse fields in which the theory of stochastic differential equations has been applied. The field that seems to rely in the most crucial way on models from the theory of SDE, is mathematical finance. However, SDE models are used in an essential way in fields ranging from, for example, bioinformatics to general relativity. Some such examples of usage of SDE models are sketched in this chapter.

## 3.1 Fatigue cracking

In *Probabilistic fracture mechanics*, one might want to consider the evolution of the length, $\ell_t$, of the largest crack in a solid material, to be governed by a stochastic differential equation. An example of this is the following model suggested by Sobczyk [54]:

$$d\ell_t = m^2 c g(Q) S^{2p} \ell_t^p \, dt + mcg(Q) S^{2p} \ell_t^p \circ dW_t.$$

In this equation, $m$ is an average of external influences, such as varying temperature and pressure, $c$ is a material parameter, $S$ is the range of the stress intensity affecting the solid, $Q$ is the stress ratio and $g$ is some specified function.

When $p = 1$, the above equation is linear and has a solution in closed form. Often it is found that $p > 1$, in which case no closed solution exists. In this case, the solution may explode in a finite time, $\tau$, corresponding to disassociation of the solid. A numerical solution of the SDE requires the simulation of trajectories and the corresponding study of approximations of $\tau$.

## 3.2 Finance

**Option pricing** Consider a simple model of a financial market consisting of only one type of stock and one type of bond. Let $S(t)$ be the price of one unit of stock, and $B(t)$ that of one bond, at time $t$. The bond is considered to be risk-free investment, and is modelled to have constant interest rate, $r$, over the time period of our study. The stock on the other hand is a risky asset. The price processes of the bond and the stock can be modelled by

$$\begin{aligned} dS(t) &= a(t, S(t))\, dt + b(t, S(t))\, dW_t \\ dB(t) &= rB(t)\, dt \end{aligned}$$

where $W$ is the one dimensional Wiener process, as before.

Into the market, described by the above equation, is now introduced a financial instrument in the form of a simple European option on the stock price, at some specified time $T$. This option is a note that gives its holder the right, but not the obligation, to by one unit of stock at the time of maturation, $T$, to a fixed price, $K$. Thus, the option pays $\Phi(T)$ on maturation, where the function $\Phi$ is defined as

$$\Phi(T) = \begin{cases} S(T) - K, & \text{if } S(T) > K \\ 0, & \text{if } S(T) \leq K \end{cases}$$

Since options are only dealt with before maturation, it is important to be able to set a reasonable price on the option. The price can be expressed as an *expectation*. This can be computed exactly only for simple models, like the famous *Black-Scholes model*. For more complicated and perhaps more realistic models, only numerical solutions are available. Thus it is important to be able to compute the expectation of the solution of the stochastic differential equation that models the evolution of the price of the underlying stock. This, and more, is developed in considerable detail in for example the book by Björk [4].

**Interest rate dynamics** Consider the so called Cox-Ingersoll-Ross model of the dynamics of interest rates, $r_t$:

$$dr_t = \alpha(\beta - r_t)\, dt + \sigma\sqrt{r_t}\, dW_t,$$

where $\alpha$, $\beta$ and $\sigma$ are positive parameters, as is the initial interest rate $r_0$. The objective is to obtain estimates of $\alpha$ and $\beta$. Using the *quasi-likelihood*

*method*, as proposed in e.g. Heyde [19], the estimates are

$$\hat{\alpha} = \frac{I_T T - J_T(r_T - r_0)}{J_t K_T - T^2}$$
$$\hat{\beta} = \frac{I_T K_T - T(r_T - r_0)}{I_T T - J_T(r_T - r_0)}$$

where, using an observed realisation of the interest rate $r_t$ over the time period $[0, T]$,

$$I_T = \int_0^T \frac{1}{r_t}\, dr_t$$
$$J_T = \int_0^T \frac{1}{r_t}\, dt$$
$$K_T = \int_0^T r_t\, dt$$

These above estimators can be evaluated using the bootstrap technique. In order to do this, it is necessary to simulate strong approximations of the integrals $I_T$, $J_T$ and $K_T$.

As a note regarding the quasi-likelihood method it can be mentioned that the estimates of $\alpha$ and $\beta$ are unaffected if the constant parameter $\sigma$ is replaced by a function of $\alpha$. See Heyde [19] pp. 134 for more on this.

## 3.3 Optimisation

The problem is that of finding the global minimum of a function

$$V : \mathbb{R}^d \to \mathbb{R},$$

without the imposition of constraints. The object of study is the equation

$$\dot{\mathbf{x}} = -\nabla V(\mathbf{x}), \tag{3.1}$$

for which the sought solution satisfies

$$\dot{\mathbf{x}} = 0.$$

A non-stochastic approach to the numerical solution of the optimisation problem has the disadvantage of "getting stuck" at a *local* minimum, instead of the global minimum. In order to improve the method, a stochastic approach has been devised. It is called *Simulated annealing*.

Instead of studying the equation (3.1), the following stochastic differential equation is studied:

$$d\mathbf{X}_t = -\nabla V(\mathbf{X}_t)\, dt + \sigma(t)\, d\mathbf{W}_t. \tag{3.2}$$

An interpretation of this equation can be given in a three dimensional setting as follows:

If we are studying the motion of a particle on a surface, we want to find the position of the lowest potential. The particle moves in a random fashion on the surface and if it gets stuck in a local minimum of the potential, $V$, then it will not stay there, because of the random motion. (One can visualise this situation as if the particle was "cooking" at the local minimum.) Thus, if the random motion is sufficiently wild, the particle will jump out of the local minimum-well and continue towards another local minimum. If the particle arrives at the *global* minimum, we do not want the particle to jump out of *this* potential-well, so the random motion must not be too wild in order for this optimisation method to work. Thus we have to make restrictions on the size of the coefficient function $\sigma(t)$. If

$$\sigma(t) = \frac{c}{\sqrt{\log(2+t)}}, \quad c > 0,$$

and the potential function, $V$, satisfies

$$\exists K > 0,\, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbf{d}} : \quad \begin{aligned} |V(\mathbf{x}) - V(\mathbf{y})| &\leq K|\mathbf{x} - \mathbf{y}| \\ |\nabla V(\mathbf{x})|^2 &\leq K(1 + |\mathbf{x}|^2) \end{aligned}$$

then, as $t \to \infty$, the *distribution* of the solution $\mathbf{X}_t$ to the stochastic differential equation (3.2) converges to a limiting distribution that has the probability mass concentrated at the point $\mathbf{x}^*$, where $\mathbf{x}^*$ is the sought global minimum of $V$. This is discussed in, for example, [29].

## 3.4 Bioinformatics

**Protein kinetics** Consider a protein that can occur in one of two forms: A and B. The structure of the protein changes over time between these two possible forms. The occurrence of one form will effect the occurrence of the other form, neccesitating the introduction of an interaction effect in a model of the dynamics of the protein. If $X_t$ is the proportion of proteins of form A at time $t$, a model for the dynamics may look like this, [29]:

$$dX_t = (a - X_t + \lambda X_t(1 - X_t))\, dt + \sigma X_t(1 - X_t) \circ dW_t.$$

Here $a$, $\lambda$ and $\sigma$ are constants. (Note that the Stratonovich integral is used.) There does not exist a closed formula for the solution, so one has to resort to numerical solutions. In this case it is interesting to obtain approximations of moments to $X_t$.

In [46], A. Neumaier uses stochastic differential equations to model the folding mechanism of proteins that give rise to their so called tertiary structure. The problem is that of determining the structure of the protein based on the sequence of amino acids comprising the protein.

## 3.5 General relativity

The famous Einstein equation in general relativity relates the dynamics of the space-time metric with the dynamics of matter-energy. The classical equation is deterministic in nature. This holds on the macroscopic scale. However, when studied in the microscopic scale, more precisely on the Planck scale ($10^{-24}$), the fluctuations of the space-time metric are assumed to be random. In order to model this, stochastic differential equations are proposed that generalise Einstein's original equation. It is observed that the pathological properties arising from using non-stochastic gravity can be removed in a stochastic gravity theory. The stochastic differential equation used is a Langevin-type equation. This is presented by Miller and Miller in [39].

# Chapter 4

# Strong solutions

The Wiener process is the stochastic driving mechanism in the stochastic differential equations considered in this thesis. As haves been mentioned, the Wiener process models *small* stochastic changes, that affect a given system. The mathematical machinery behind all this is introduced in the following sequence of definitions.

**Definition 4.0.1** *A* filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$*, is a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, equipped with a family of* $\sigma$*-algebras* $\{\mathcal{F}_t\}_{t\geq 0}$*, called a* filtration*, such that the following properties hold:*

1. $\forall 0 \leq s < t < \infty : \ \mathcal{F}_s \subset \mathcal{F}_t$;
2. $\mathcal{F}_\infty \equiv \sigma(\bigcup_{t\geq 0} \mathcal{F}_t) \subset \mathcal{F}$.

**Definition 4.0.2** *A* right-continuous *filtered probability space, is a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$*, such that*

$$\forall t \geq 0 : \ \mathcal{F}_t = \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}.$$

**Definition 4.0.3** *An* augmented *filtered probability space, is a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$*, such that* $\mathcal{F}_0$ *contains all* $\mathbb{P}$*-null sets.*

**Definition 4.0.4** *A stochastic process,* $\{X_t\}_{t\geq 0}$*, defined on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$*, is* adapted *if*

$$\forall\, t \geq 0 : \quad X_t \in \mathcal{F}_t.$$

**Definition 4.0.5** *A (standard)* Wiener process*, is a continuous and adapted stochastic process* $W = \{W_t\}_{t\geq 0} = \{W(t)\}_{t\geq 0}$*, defined on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$*, such that the following conditions hold:*

1. $\forall s \in [0,t)$ : *$W_t - W_s$ is independent of the $\sigma$-algebra $\mathcal{F}_s$;*

2. $\forall s \in [0,t)$ : *$W_t - W_s$ is a normally distributed random variable with mean $\mathbb{E}[W_t - W_s] = 0$ and variance $\mathrm{Var}[W_t - W_s] = t - s$;*

3. $\mathbb{P}(W_0 = 0) = 1$.

## 4.1 Definition, existence and uniqueness

Consider the one dimensional stochastic differential equation

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dW_t. \tag{4.1}$$

As have been mentioned already, there is a concept of a strong solution associated to this equation:

**Definition 4.1.1** *Fix a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, together with a Wiener process, $W$, on that probability space. Let $X = \{X_t\}_{t\geq 0}$ be a stochastic process with the following properties:*

1. *the sample paths of $X$ are continuous;*

2. *$X$ is adapted;*

3. $\forall t \in [0,\infty)$ : $\mathbb{P}\left( \int_0^t |a(s, X_s)|\, ds + \int_0^t b^2(s, X_s)\, ds < \infty \right) = 1$;

4. $\mathbb{P}\left( \forall t \in [0,\infty) : X_t = X_0 + \int_0^t a(s, X_s)\, ds + \int_0^t b(s, X_s)\, dW_s \right) = 1$.

*Then the process $X$ is called a* strong solution *to the stochastic differential equation* (4.1).

If the equation (4.1) is thought of as representing a dynamical system that has the Wiener process, $W$, as input and response $X$ as output, then the requirement that $X$ is adapted to the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is merely the requirement that the system obeys the *principle of causality*.

Having stated the definition of a strong solution, it is important to state in what sense a strong solution is unique, if it exists:

**Definition 4.1.2** *If, for any two strong solutions $X$ and $Y$ of the stochastic differential equation* (4.1)*, it holds that*

$$\mathbb{P}\Big(\forall t \in [0,\infty) : X_t = Y_t\Big) = 1,$$

*then the equation is said to have a* unique *strong solution.*

**Remark 4.1.1** A pair of stochastic processes $X = \{X_t\}_{t\geq 0}$ and $Y = \{Y_t\}_{t\geq 0}$ satisfying

$$\mathbb{P}\Big(\forall t \in [0,\infty) : X_t = Y_t\Big) = 1$$

are said to be *indistinguishable.* This means that their respective sample paths are the same, almost surely.

Having defined the concept strong solutions, and the sense in which such solutions are unique, we now come to the task of giving conditions for the existence of a strong solution. This is accomplished in the following theorem of Yamada and Watanabe in [64]:

**Theorem 4.1.1** *Let the coefficients of the one dimensional stochastic differential equation* (4.1) *satisfy the following conditions:*

1. $\forall t \in [0,\infty),\ \forall x, y \in \mathbb{R} : \quad |a(t,x) - a(t,y)| \leq \kappa(|x-y|);$
2. $\forall t \in [0,\infty),\ \forall x, y \in \mathbb{R} : \quad |b(t,x) - b(t,y)| \leq \sqrt{h(|x-y|)},$

*where the functions $\kappa, h : [0,\infty) \to [0,\infty)$ are such that:*

3. $\kappa(0) = 0$ *and* $h(0) = 0;$
4. $\kappa$ *and* $h$ *are strictly increasing;*
5. $\kappa$ *is concave;*
6. $\forall \epsilon > 0 : \quad \displaystyle\int_0^\epsilon \frac{dt}{\kappa(t)} = \infty;$
7. $\forall \epsilon > 0 : \quad \displaystyle\int_0^\epsilon \frac{dt}{h(t)} = \infty.$

*Then the equation possesses a unique strong solution.*

Since most stochastic differential equations do not have a strong solution expressible in closed form, it is necessary to be able to obtain numerical approximations. In the upcoming chapter, several different numerical methods will be presented that give *strong approximations* to a stochastic differential equation.

# Chapter 5

# Strong approximations

This chapter occupies a considerable portion of the thesis, since there is a lot of work published on strong solutions of stochastic differential equations. One of the main reasons for this in turn, is that numerical approximation of strong solutions take their inspiration from the well developed theory of numerical solution of ordinary differential equations.

In this chapter, numerical methods that give *strong approximations* to stochastic differential equations will be addressed. These methods will involve concepts from differential geometry and concepts from complex variable theory in several dimensions.

Before going into this, it is necessary to define what is meant by a strong approximation. For this, it is suitable to first consider an example of a one dimensional stochastic differential equation. Let that equation be

$$\begin{aligned} dX_t &= a\,X_t\,dt + b\,X_t\,dW_t \\ X_0 &= 1 \\ t &\in [0,1] \end{aligned} \tag{5.1}$$

An approximation, $\{Y_k^n\}_{k=0}^n$, to the solution, $\{X_t\}_{t\in[0,1]}$, of the stochastic differential equation (5.1), is a *time-discrete* stochastic process, such that $Y_k^n$ is an approximation to $X(t_k^n)$. Here, $0 = t_0^n < t_1^n < \ldots t_n^n = 1$, is a partition of $[0,1]$.

The $Y_k^n$:s are computed according to some scheme that is connected to the equation (5.1). For example, they may be computed by the recursive scheme

$$\begin{aligned} Y_{k+1}^n &= Y_k^n + aY_k^n h_k^n + bY_k^n \Delta W_k^n \\ Y_0 &= X_0 = 1 \end{aligned} \tag{5.2}$$

where $h_k^n = t_{k+1}^n - t_k^n$ and $\Delta W_k = W_{t_{k+1}} - W_{t_k}$. The *step sizes*, $h_k^n$, satisfy

$$\lim_{n\to\infty} \max_{1\le k\le n} h_k^n = 0.$$

Since the approximation error accumulates during the iterations, a measure of the accuracy of the numerical approximation can be obtained by comparing the final value of the approximation, $Y_n^n$, with the final value of the exact solution, $X(1)$. The process $Y$ is said to be a *strong approximation* to $X$, or alternatively is said to *converge strongly* to $X$, if

$$\lim_{n\to\infty} \mathbb{E}\big[|X(1) - Y_n^n|\big] = 0. \tag{5.3}$$

The approximation *converges strongly with order* $p$ if

$$\exists C > 0,\, \exists N \geq 1,\, \forall n > N: \quad \mathbb{E}\big[|X(1) - Y_n^n|\big] \leq C\, h_n^p,$$

where

$$h_n = \max_{1\leq k\leq n} h_k^n.$$

If this holds, it is customary to just say that the scheme (5.2) is of strong order $p$.

**Remark 5.0.2** Notice that, with the above language, it is a matter of $\mathbb{L}^1$-convergence, rather than strong convergence in the usual sense convergence with probability 1.

## 5.1 Stochastic Taylor expansions

The technique of stochastic Taylor expansions was proposed by Platen and Wagner in [48], and a thorough discussion of this topic can be found in [29]. Here the basic idea is presented.

Until further notice, $X_t$ denotes the solution to the one dimensional stochastic differential equation

$$dX_t = a(X_t)\, dt + b(X_t)\, dW_t. \tag{5.4}$$

Notice that the coefficient $a$ and $b$ do not depend explicitly on $t$. Such equations are called *autonomous* (or time homogeneous).

If the functions $a$ and $b$ are sufficiently regular, so as to allow the application of the Itô formula (2.6), then the following representations of $a(X_t)$ and $b(X_t)$ hold:

$$a(X_t) = a(X_{t_0}) + \int_{t_0}^t L_0 a(X_s)\, ds + \int_{t_0}^t L_1 a(X_s)\, dW_s$$
$$b(X_t) = b(X_{t_0}) + \int_{t_0}^t L_0 b(X_s)\, ds + \int_{t_0}^t L_1 b(X_s)\, dW_s$$

Here the operators $L_0$ and $L_1$ are given by

$$\begin{aligned} L_0 &= a\,\frac{d}{dx} + \frac{1}{2}\,b^2\frac{d^2}{dx^2} \\ L_1 &= b\,\frac{d}{dx} \end{aligned} \tag{5.5}$$

**Remark 5.1.1** The second term in the operator $L_0$ comes from using the Itô integral to interpret the stochastic differential equation. If instead one uses the Stratonovich integral, then the operator $L_0$ will be replaced with

$$\tilde{L_0} = a\,\frac{d}{dx},$$

while

$$\tilde{L_1} = L_1 = b\,\frac{d}{dx},$$

as before.

Insert the Itô representations of $a(X_t)$ and $b(X_t)$ into the integral representation of the solution to (5.4)

$$X_t = X_{t_0} + \int_{t_0}^{t} a(X_{s_1})\,ds_1 + \int_{t_0}^{t} b(X_{s_1})\,dW_{s_1}.$$

Then we get

$$\begin{aligned} X_t &= X_{t_0} + a(X_{t_0})\int_{t_0}^{t} ds_1 + b(X_{t_0})\int_{t_0}^{t} dW_{s_1} \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s_1} L_0 a(X_{s_2})\,ds_2 ds_1 + \int_{t_0}^{t}\int_{t_0}^{s_1} L_1 a(X_{s_2})\,dW_{s_2} ds_1 \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s_1} L_0 b(X_{s_2})\,ds_2 dW_{s_1} + \int_{t_0}^{t}\int_{t_0}^{s_1} L_1 b(X_{s_2})\,dW_{s_2} dW_{s_1} \\ &= X_{t_0} + a(X_{t_0})\int_{t_0}^{t} ds_1 + b(X_{t_0})\int_{t_0}^{t} dW_{s_1} + R_1, \end{aligned}$$

where the remainder $R_1$ is given by

$$\begin{aligned} R_1 &= \int_{t_0}^{t}\int_{t_0}^{s_1} L_0 a(X_{s_2})\,ds_2 ds_1 + \int_{t_0}^{t}\int_{t_0}^{s_1} L_1 a(X_{s_2})\,dW_{s_2} ds_1 \\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s_1} L_0 b(X_{s_2})\,ds_2 dW_{s_1} + \int_{t_0}^{t}\int_{t_0}^{s_1} L_1 b(X_{s_2})\,dW_{s_2} dW_{s_1}. \end{aligned}$$

Since the functions $a(X_{s_2})$ and $b(X_{s_2})$ can be represented in the same way as $a(X_{s_1})$ and $b(X_{s_1})$, this procedure gives a representation of $X_t$ in terms of a weighted sum of *multiple Itô integrals*:

$$\begin{aligned} X_t = X_{t_0} &+ a(X_{t_0}) \int_{t_0}^{t} ds_1 + b(X_{t_0}) \int_{t_0}^{t} dW_{s_1} \\ &+ L_0 a(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s_1} ds_2 ds_1 + L_1 a(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s_1} dW_{s_2} ds_1 \\ &+ L_0 b(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s_1} ds_2 dW_{s_1} + L_1 b(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s_1} dW_{s_2} dW_{s_1} \\ &+ R_2. \end{aligned} \tag{5.6}$$

The remainder, $R_2$, is given by the following rather involved expression:

$$\begin{aligned} R_2 = &\int_{t_0}^{t} \int_{t_0}^{s_1} L_0 a(X_{s_2})\, ds_2 ds_1 + \int_{t_0}^{t} \int_{t_0}^{s_1} L_1 a(X_{s_2})\, dW_{s_2} ds_1 \\ &+ \int_{t_0}^{t} \int_{t_0}^{s_1} L_0 b(X_{s_2})\, ds_2 dW_{s_1} + \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{t_0}^{s_2} L_0 L_1 b(X_{s_3})\, ds_3 dW_{s_2} dW_{s_1} \\ &+ \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{t_0}^{s_2} L_1 L_1 b(X_{s_3})\, dW_{s_3} dW_{s_2} dW_{s_1}, \end{aligned}$$

where

$$L_i a(X_{t_0}) = L_i a(X_t)\big|_{t=t_0}, \quad i = 0, 1.$$

An expansion of the above kind, using Itô integrals, is called an *Itô-Taylor expansion*. The *Stratonovich-Taylor expansion* is analogous, except that it uses Stratonovich stochastic integrals instead of Itô integrals.

## 5.2 The Euler-Maruyama method

The simplest numerical approximation scheme, that can be derived using the stochastic Taylor expansion, is a stochastic counterpart of the well known Euler method from the numerical solution of ordinary differential equations.

Only the first two terms are retained in the stochastic Taylor expansion (5.6), of the solution to the stochastic differential equation. This gives the approximation

$$X_t \approx X_{t_0} + a(X_{t_0}) \int_{t_0}^{t} ds_1 + b(X_{t_0}) \int_{t_0}^{t} dW_{s_1}.$$

To use the above approximation to construct a numerical scheme, first partition the interval $[t_0, t]$ uniformly:

$$\begin{aligned} t_0 = \tau_0 < \tau_1 < \cdots < \tau_n = t \\ h = \tau_{k+1} - \tau_k \end{aligned} \tag{5.7}$$

Then construct a numerical scheme, which will give approximate values to the exact solution values, $X_{\tau_k}$, at the nodes, $\tau_k$, in the partition, in the following way: Let the stochastic process, $\overline{X} = \{\overline{X}_k\}_{k=0}^n$, defined by

$$\overline{X}_k = X_{\tau_k},$$

satisfy the equation

$$\overline{X}_{k+1} = \overline{X}_k + a(\overline{X}_k)\, h + b(\overline{X}_k)\, \Delta_k W, \tag{5.8}$$

where

$$\Delta_k W = W_{\tau_{k+1}} - W_{\tau_k}.$$

The process $\overline{X}$ is called the *Euler-Maruyama approximation*, of the solution to the equation (5.4), and the numerical scheme (5.8) is the *Euler-Maruyama scheme*, to compute that approximation. It was suggested by Maruyama [37].

The process $\overline{X}$ is a Markov process, by construction, and the solution to the stochastic differential equation is a diffusion process. So what we are doing here, in essence, is to approximate a continuous-time diffusion with a discrete time Markov process. This Markov process does not have finite state space. However, in [30], Kushner approximated the solution of the stochastic differential equation by a Markov chain, thus discretising both the time variable $t$ and the values of the solution $X$.

In [29], Kloeden and Platen prove that the Euler-Maruyama scheme is a numerical approximation scheme of strong order $\frac{1}{2}$, i.e.

$$\exists C > 0,\ \exists N \geq 1,\ \forall n > N: \quad \mathbb{E}\big[|X_1 - \overline{X}_n|\big] \leq C\, h^{1/2}.$$

From a theoretical perspective, the Euler-Maruyama scheme is a good choice, since it is quite amenable to investigations regarding the *stability*, *rate of convergence* and *consistency*. From a computational point of view, however, it is a bad choice, since it requires a very fine partition of the time interval, in order to give accurate approximations.

It is worth emphasising that, when talking about the Euler-Maruyama scheme, or any other numerical scheme that approximates the differentiation

operator, nothing is said as to the nature of the solution process being weak or strong.

Any numerical scheme focuses on approximating the solution of the differential equation. If a *strong solution* is sought, then the following issues have to be addressed:

- How is the driving process in a stochastic differential equation to be represented numerically?
- How should the time interval be partitioned?

If a *weak solution* is sought, then it might not be necessary to simulate the driving process as it stands, but rather replace it with a *simpler process*, that possesses some *vital characteristics* of the original driving process, for example its first two moments. The reason for this simplification may be to obtain a model more amenable to theoretical investigations, or to gain in computational speed.

## 5.3 The Milstein method

If the first, second and sixth integrals in the expansion (5.6) are retained, then the *Milstein approximation* is obtained:

$$\begin{aligned} X_t \approx X_{t_0} + a(X_{t_0}) \int_{t_0}^{t} ds_1 + b(X_{t_0}) \int_{t_0}^{t} dW_{s_1} \\ + L_1 b(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s_1} dW_{s_2} dW_{s_1}. \end{aligned} \tag{5.9}$$

Here the coefficient $L_1 b(X_{t_0})$ is, according to the definition (5.5) of the operator $L_1$, given by

$$L_1 b(X_{t_0}) = b(x) \frac{db(x)}{dx}\Big|_{x=X_{t_0}}.$$

The integrals that feature in (5.9) are approximated with appropriate sums. This leads us to a central issue in the business of stochastic Taylor expansions, namely that of approximating so called *multiple stochastic integrals*. Of such,

$$\int_{t_0}^{t} \int_{t_0}^{s_1} dW_{s_2} dW_{s_1}$$

is the simplest non-trivial example. For the approximation, the following result from Karatzas and Shreve [28] is important: Recall that a Hermite

polynomial, $H_n$, of order $n$, can be defined as

$$H_n(x) = (-1)^n \, \mathrm{e}^{x^2} \frac{d^n}{dx^n}\big(\mathrm{e}^{-x^2}\big).$$

**Theorem 5.3.1** *Let $M$ be a continuous local martingale and let $H_n$ be the Hermite polynomial of order $n$. Then we have*

$$\int_0^t \int_0^{s_1} \dots \int_0^{s_{n-1}} dM_{s_n} \dots dM_{s_2} dM_{s_1} = \frac{1}{n!} \mathcal{H}_n\Big(M_t, [M,M]_t\Big),$$

*where*

$$\mathcal{H}_n(x,y) = \sqrt{n!}\, y^{n/2} H_n\Big(\frac{x}{\sqrt{y}}\Big),$$

*and $[M,M]$ is the quadratic variation process of $M$.*

Thus, multiple Wiener-Itô integrals can be expressed using Hermite polynomials. For definitions and properties of local martingales, and of the quadratic variation process, see for example the book by Karatzas and Shreve [28].

Since the Wiener process, $W$, is a continuous local martingale with quadratic variation

$$[W,W]_t = t,$$

the above theorem can be used to compute, for example

$$\int_0^t \int_0^{s_1} dW_{s_2} dW_{s_1} = \frac{1}{2}\mathcal{H}_2\big(W_t, t\big) = \frac{1}{2}\big(W_t^2 - t\big),$$

where we used that

$$\mathcal{H}_2(x,y) = x^2 - y.$$

Thus, the discrete version of the Milstein approximation to the solution process, $X$, is a stochastic process, $Y$, satisfying:

$$\begin{aligned} Y_{k+1} &= Y_k + a(Y_k)\, h + b(Y_k)\, \Delta_k W + \frac{1}{2} b(Y_k) b'(Y_k) \left((\Delta_k W)^2 - h\right) \\ Y_k &= X_{\tau_k} \end{aligned} \tag{5.10}$$

In [29], Kloeden and Platen prove that this scheme is of strong order 1.

## 5.4 Simulation of iterated Itô integrals

The fundamental building blocks in a stochastic Taylor expansion are the multiple stochastic integrals, amongst which

$$\int_{t_0}^{t} dW_{s_1}, \quad \int_{t_0}^{t}\int_{t_0}^{s_1} dW_{s_2} ds_1, \quad \int_{t_0}^{t}\int_{t_0}^{s_1} ds_2 dW_{s_1} \quad \text{and} \quad \int_{t_0}^{t}\int_{t_0}^{s_1} dW_{s_2} dW_{s_1}$$

are examples. The purpose of a stochastic Taylor expansion, other than having an alternative representation of the solution to the stochastic differential equation, is to obtain approximate solutions. This is used when finding strong numerical solutions to a given stochastic differential equation. For this, it is necessary to be able to simulate multiple stochastic integrals.

### 5.4.1 A single Wiener process

A general multiple stochastic integral of the Wiener process, using the Itô interpretation of the stochastic integral, can be written in the form:

$$I_{i_1,\ldots,i_n}(t) = \int_0^t \int_0^{s_1} \ldots \int_0^{s_{n-1}} dW_{i_n}(s_n) \ldots dW_{i_2}(s_2) dW_{i_1}(s_1). \tag{5.11}$$

Here we use the notation:

$$dW_i(s) = \begin{cases} ds & i = 0 \\ dW_s & i = 1 \end{cases}$$

The multiple integral (5.11) is simulated by means of recursion:

$$I_{i_1,\ldots,i_n}(t) = \int_0^t I_{i_1,\ldots,i_{n-1}}(s_n)\, dW_{i_n}(s_n).$$

As an illustration of this procedure, consider the simulation of the iterated integral

$$I_{101}(t) = \int_0^t \int_0^{s_1} \int_0^{s_2} dW_1(s_3) dW_0(s_2) dW_1(s_1).$$

By recursion, the integral is broken up as follows:

$$\begin{aligned} I_{101}(t) &= \int_0^t I_{10}(s_3)\, dW_1(s_3) = \int_0^t I_{10}(s_3)\, dW_{s_3} \\ I_{10}(s_3) &= \int_0^{s_3} I_1(s_2)\, dW_0(s_2) = \int_0^{s_3} I_1(s_2)\, ds_2 \\ I_1(s_2) &= \int_0^{s_2} dW_1(s_1) \qquad = \int_0^{s_2} dW_{s_1} \qquad = W_{s_2} \end{aligned}$$

Reassembling the parts, we get

$$I_{101}(t) = \int_0^t I_{10}(s_3)\, dW_{s_3} = \int_0^t \left( \int_0^{s_3} W_{s_2}\, ds_2 \right) dW_{s_3}$$
$$= W_t \int_0^t W_{s_2}\, ds_2 - \int_0^t W_{s_2}^2\, ds_2$$

For the last step, partial integration was used:

$$\int u\, dv = uv - \int v\, du.$$

This reduces the simulation of the multiple integral

$$\int_0^t \int_0^{s_1} \int_0^{s_2} dW_{s_3} ds_2 dW_{s_1},$$

to that of simulating the integrals

$$\int_0^t W_{s_2}\, ds_2 \quad \text{and} \quad \int_0^t W_{s_2}^2\, ds_2.$$

### 5.4.2 Multiple Wiener processes

Stochastic differential equations involving several Wiener processes are considerably more difficult to obtain approximate solutions for, due to the complexity of simulating the iterated integrals occurring in the stochastic Taylor expansion. In [50], Ryden and Wiktorsson use a result of Lévy [33] to devise a way to approximate the integral

$$\int_0^t \int_0^{s_1} dW_1(s_2) dW_2(s_1),$$

where $W_1$ and $W_2$ are independent Wiener processes.

We will not pursue the the topic of stochastic Taylor expansions, and the related techniques of numerically approximating multiple stochastic integrals, except to point the interested reader to some literature on the subject.

For an exhaustive discussion of stochastic Taylor expansions for SDE:s of one and several variables, the standard work on numerical solution of stochastic differential equations [29] by Kloeden and Platen is the best source. They do not however discuss the numerical approximation of multiple stochastic integrals. The book [41] by Milstein does this, both for the more simple case of only one driving Wiener process, and also to some extent for the more complicated case of several driving Wiener processes.

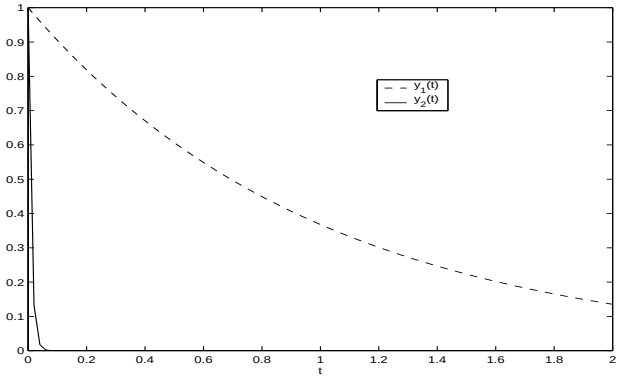
## 5.5 Waveform relaxation

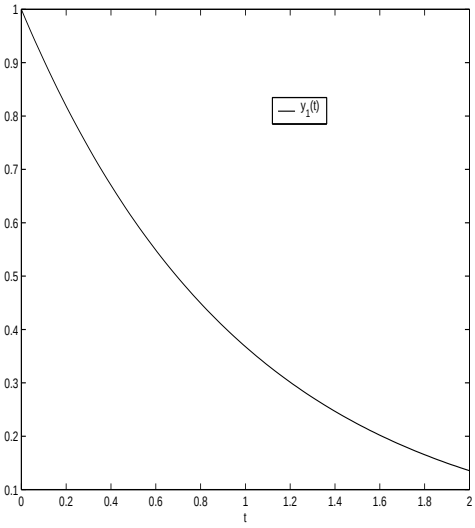
### 5.5.1 Application to ordinary differential equations

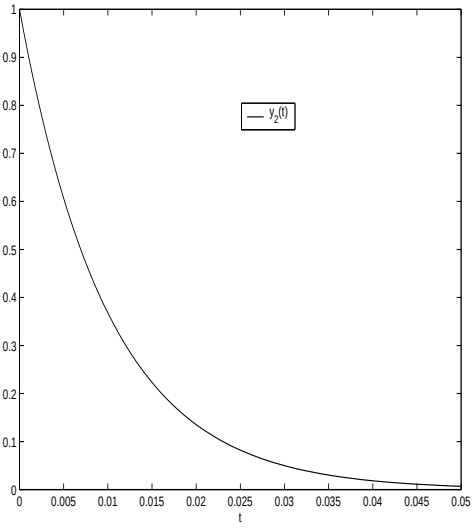
Consider the following system of ordinary differential equations, where each component has its own time scale.

$$\frac{dy_1}{dt} = -y_1(t)$$
$$\frac{dy_2}{dt} = -100y_2(t)$$

As is seen in the following figures, the time scales of the components differ considerably.




This is the typical behaviour of a *stiff* system of differential equations. When solving this system numerically, one has to use a time step that is small enough to capture the quick variations of $y_2(t)$. For $y_1(t)$ this is a waste of

computational resources, since it varies rather slowly compared with $y_2(t)$. Since the system can be broken up into two separate, non-interacting sub-systems

$$\frac{dy_1}{dt} = -y_1(t) \quad \text{and} \quad \frac{dy_2}{dt} = -100y_2(t),$$

it would be worthwhile to have a numerical method that solves the system in parallel, and to use a time step that is individually adapted to each of the components of the solution vector $y(t) = (y_1(t), y_2(t))$.

The above example is very simple, but still gives a flavour of a more realistic one, like for example computing a solution vector of a system of ordinary differential equations, consisting of as many as perhaps a thousand equations.

Using universally a time step that resolves the fastest varying component is grossly inefficient. Certainly, if applicable, a parallel numerical solution method, along the lines indicated above, would be worthwhile. One such family of method is the *Waveform relaxation methods*. They are used on very large systems of *weakly interacting* differential equations. The weak interaction enables the whole system to be broken down into non-interacting sub-systems, which can then be solved in parallel.

The problem of numerically solving large systems of differential equations is typical in the numerical simulation of VLSI (Very Large Scale Integrated) circuits, and indeed it was in this area that the waveform relaxation method was developed by Lelarasmee in his PhD thesis [32]. A presentation of waveform relaxation methods can be found in the book [59] by Vandewalle.

**Features of a waveform relaxation method** As an example of the way the waveform relaxation method works, consider the system of equations:

$$\begin{aligned} \frac{dx}{dt} &= y, \quad x(0) = 0 \\ \frac{dy}{dt} &= -x, \quad y(0) = 1 \end{aligned}$$

The solution is of course

$$\begin{aligned} x(t) &= \sin(t) \\ y(t) &= \cos(t) \end{aligned}$$

A Gauss-Seidel waveform relaxation scheme for this problem is

$$\begin{cases} x_1'(t) = y_0(t) \\ y_1'(t) = -x_1(t) \end{cases} \Rightarrow \begin{cases} x_1(t) &= t \\ y_1(t) &= 1 - \frac{t^2}{2!} \end{cases}$$

$$\begin{cases} x_2'(t) = y_1(t) \\ y_2'(t) = -x_2(t) \end{cases} \Rightarrow \begin{cases} x_2(t) &= t - \frac{t^3}{3!} \\ y_2(t) &= 1 - \frac{t^2}{2!} + \frac{t^4}{4!} \end{cases}$$
$$\vdots$$

The pair $(x_n(t), y_n(t))$ is the approximate solutions after the n:th iteration. Note that the scheme gives the Taylor expansions of $\sin(t)$ and $\cos(t)$.

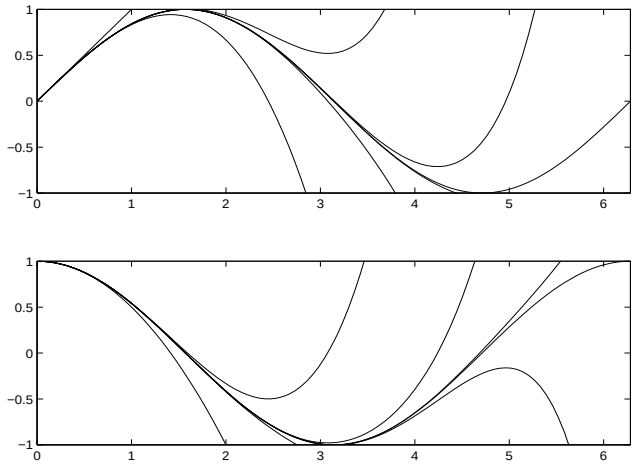


Figure 5.1: Successive iterates for the Gauss-Seidel waveform relaxation scheme.

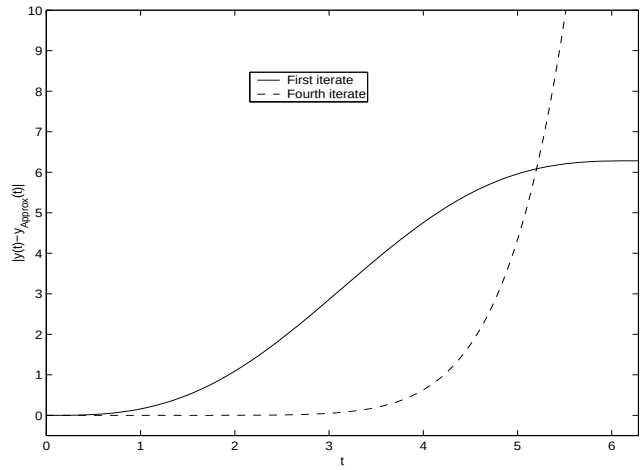



Figure 5.2: Error of the first and fourth waveform relaxation iterate for $y(t)$.

The features, illustrated in Figure 5.2 above, are typical for any waveform relaxation scheme, namely that:

- The error of a specific iterate is small in the beginning of the interval $[0, 2\pi]$, but grows further in.

- Each iteration lengthens the interval over which the error is small. item The error of successive iterates is not necessarily uniformly decreasing: At the end of the interval $[0, 2\pi]$, the error of the fourth iterate is larger than the error of the first iterate. This fact indicates that the convergence rate is very slow for a waveform relaxation method.

Due to the slow convergence rate, some of the current research is focused on accelerating the convergence. Leimkühler [31] and Janssen [26] are examples of the articles that address this issue.

**General formulation of the waveform relaxation method** Many phenomena that are modelled as ordinary differential equations can be cast in the following general form: Find the function $x$ that satisfies the equation

$$x = T(x) + g, \tag{5.12}$$

where $x \in U$, $g \in U$ and $T : U \to U$ is an operator from the function space $U$ into $U$.

If the equation (5.12) represents a (large) system of *weakly* interacting subsystems, then the weak interaction can be used to recast the problem of finding $x$ in a different form: Decompose the function space $U$ into a product of subspaces $\{U_j\}_{j=1}^n$:

$$U = U_1 \times \ldots \times U_n.$$

This decomposition induces a corresponding decomposition of the operator $T$ into $\{T_j\}_{j=1}^n$, and the function $g$ into $\{g_j\}_{j=1}^n$. The equation (5.12) now takes the form:

$$\begin{aligned} x^{(1)} &= T_1(x^{(1)}, \ldots, x^{(n)}) + g_1 \\ &\vdots \\ x^{(n)} &= T_n(x^{(1)}, \ldots, x^{(n)}) + g_n \end{aligned} \tag{5.13}$$

where $x^{(j)}$, $g_j \in U_j$ and $T_j : U \to U_j$. Each subsystem

$$x^{(j)} = T_j(x^{(1)}, \ldots, x^{(n)}) + g_j$$

is solved, and the solutions are combined to give the solution of the whole system. The decomposition is the starting point for iteration schemes, e.g.

the Gauss-Seidel scheme:

$$\begin{aligned} x_i^{(1)} &= T_1(x_{i-1}^{(1)}, x_{i-1}^{(2)} \ldots, x_{i-1}^{(n)}) + g_1 \\ x_i^{(2)} &= T_2(x_i^{(1)}, x_{i-1}^{(2)} \ldots, x_{i-1}^{(n)}) + g_2 \\ &\vdots \\ x_i^{(n)} &= T_n(x_i^{(1)}, x_i^{(2)} \ldots, x_i^{(n-1)}, x_{i-1}^{(n)}) + g_n \end{aligned} \tag{5.14}$$

The significant difference between waveform relaxation methods, and non-waveform relaxation methods, is that the waveform relaxation method calculates successive approximations of the *function* (or waveform) $x$, not just approximations for specific values of the function $x$, as non-waveform relaxation methods do, hence the name *waveform* relaxation method. One can say that waveform relaxation methods compute a "uniform approximation" to $x$, as opposed to non-waveform relaxation methods which compute a "point-wise approximation".

In [52], Schneider and Schurz address the question of convergence of the iterative scheme (5.13) :

**Theorem 5.5.1** *Assume that*

$$\forall j \in \{1 \ldots n\} : \ U_j \text{ is a Banach space with norm } ||\cdot||_j$$

*and*

$$\forall j \in \{1 \ldots n\} : \ T_j : U \to U_j \text{ is linear, continuous and globally Lipschitz, with } L = (l_{jk})_{1 \le j,k \le n}, \text{ the matrix of Lipschitz coefficients.}$$

*Further, assume that the spectral radius*

$$\rho(L) = \sup_{1 \le j \le n} |\lambda_j|$$

*of $L$ satisfies $\rho(L) < 1$, where $\{\lambda_j\}_{j=1}^n$ are the eigenvalues of $L$. Then the iteration scheme (5.14) converges in $U = U_1 \times \ldots \times U_n$, with respect to the following norm $||\cdot||_U$:*

$$||x||_U = \sum_{j=1}^{n} ||e_j||_j \, ||x^{(j)}||_j.$$

*Here $e = (e_1, \ldots, e_n)$ is a strictly positive eigenfunction (in the language of Schneider and Schurz [52]), corresponding to $\rho(L)$, i.e.*

$$\forall j \in \{1, \ldots, n\} : \ L(e) = \rho(L)\, e \quad \text{and} \quad ||e_1||_1, \ldots, ||e_n||_n > 0.$$

The proof of this theorem is similar in spirit to the proof of the fixed-point theorem for Banach spaces. See for example Debnath and Mikusinski [9].

### 5.5.2 Application to stochastic differential equations

The idea behind Theorem 5.5.1 can be applied to systems of SDE:s

$$\begin{aligned} dX(t) &= \sum_{k=0}^{m} f_k(t, X(t))\, dW_k(t) \\ X(0) &= X_0 \end{aligned} \tag{5.15}$$

To that end, it has to be addressed what effect the introduction of randomness will have on the analogues of the Lipschitz-coefficients, that feature in Theorem 5.5.1. Of course, the theory is more involved in a stochastic context. Nevertheless, it *is* possible to adapt waveform-relaxation methods to stochastic contexts.

To set the stage for a stochastic counterpart of theorem (5.5.1), consider an augmented filtered probability space $\mathcal{P} = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$. The first step is to split (5.15) into a system of $n$ *weakly* interacting parts:

$$\begin{cases} dX_1(t) = \sum_{k=0}^{m} f_{1k}(t, X_1(t), \ldots, X_n(t))\, dW_k(t) \\ X_1(0) = X_{10} \\ \qquad\qquad \vdots \\ dX_n(t) = \sum_{k=0}^{m} f_{nk}(t, X_1(t), \ldots, X_n(t))\, dW_k(t) \\ X_n(0) = X_{n0} \end{cases} \tag{5.16}$$

where the functions

$$\left\{ f_{jk} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d_j} \right\}_{(j,k)\,\in\,\{1,\ldots,n\}\times\{0,\ldots,m\}}, \quad d_1 + \ldots + d_n = d,$$

are assumed to be Lebesgue-measurable, and have the following properties:

- $(A_1)$ $$\forall j,k: \quad \sup_{t\in[0,T]} \inf_{y\in\mathbb{R}^d} ||f_{j,k}(t,y)||_j \leq K < \infty$$

- $(A_2)$ $$\forall i,j \in \{1,\ldots,n\},\ \forall k \in \{0,\ldots,m\},$$
$$\exists L_{i,k}^{(j)} \in \mathbb{R},\ \forall (x_1,\ldots,x_n), (y_1,\ldots,y_n) \in \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_n}\ :$$
$$||f_{jk}(t,x_1,\ldots,x_n) - f_{jk}(t,y_1,\ldots,y_n)||_j \leq \sum_{i=1}^{n} L_{ik}^{(j)} ||x_i - y_i||_j,$$

  i.e. each function $f_{jk}$ is globally Lipschitz continuous in $x$, uniformly with respect to $t$.

Given a $p \geq 1$, the waveform-relaxed solution converges to the solution of the system of stochastic differential equations (5.16), in the space $\mathcal{E}^p$ defined in the following way:

$$\mathcal{E}^p = \mathcal{E}_0 \bigcap \mathcal{E}_1^p$$

where

$$\mathcal{E}_0 = \Big\{ X : [0,T] \times \Omega \to \mathbb{R}^d : X \in \mathcal{C}([0,T]),\, X_t \in \mathcal{F}_t \Big\}$$

$$\mathcal{E}_1^p = \Big\{ X : [0,T] \times \Omega \to \mathbb{R}^d : \mathbb{E}\Big[ \max_{t \in [0,T]} ||X_t||^p \Big] < \infty \Big\}$$

Here the norm $||\cdot||$ is defined in the same way as $||\cdot||_U$ in Theorem 5.5.1.

The space $\mathcal{E}^p$ is a Banach space, with respect to the norm

$$||X||_{\mathcal{E}^p}^p = \mathbb{E}\Big[ \sup_{t \in [0,T]} ||X_t||^p \Big].$$

The following theorem, of Schneider and Schurz [52], addresses the convergence of waveform-relaxed solutions to the system of SDE:s (5.16):

**Theorem 5.5.2** *Assume that the coefficients* $\{f_{jk}\}$ *of the system* (5.16) *are Lebesgue-measurable and satisfy the conditions* $(A_1)$ *and* $(A_2)$. *Further, assume that, for a given* $p \geq 1$,

$$\mathbb{E}\Big[ \max_{1 \leq j \leq n} ||X_{j0}||_j \Big] < \infty.$$

*Then the initial value problem* (5.16) *has a unique solution in* $\mathcal{E}^p$.

In essence, the proof is about showing that the assumptions imply the existence of a spectral radius $\rho(L) < 1$ for a matrix $L$ determined by the Lipschitz-coefficients. This makes the operator defined by the system of SDE:s (5.16) a contraction from $\mathcal{E}^p$ to $\mathcal{E}^p$.

**Perspective for waveform relaxation applied to SDE** So far, there do not seem to be much more available in terms of applications of the waveform relaxation method to stochastic differential equations. But it seems likely that there will be more research in this area, since large systems of stochastic differential equations are more realistic when modelling the dependencies between a large number of entities in, for example, financial markets.

It seems highly probable that the price of Ericsson stock, say, depends on other stock prices in the telecommunications market, as well as on other stocks. Numerical analysis of this seems to be amenable to the technique of waveform relaxation. A possible difficulty might be the requirement of the

waveform relaxation method of needing the assumption of weak interaction between the equations in the system of stochastic differential equations. In a specific context, this weak dependency first has to be established, before applying the numerical method.

## 5.6 Adaptive methods

The methods mentioned so far always assumed a *fixed* partition

$$0 = t_0 < t_1 < \ldots < t_n = T$$

of $[0, T]$, when computing the approximate solution of the SDE in question. In order to improve the approximation, the step-sizes were made smaller and smaller, resulting in increased computation time. Looking at the SDE

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dW_t,$$

one notices that the underlying driving stochastic noise, $W_t$, is coupled to the evolution of the process, $X_t$, through the diffusion coefficient $b$.

In [20], Hofmann, Müller-Grönbach and Ritter suggest that, if the coefficient $b$ is big, then the coupling will be strong, indicating that a small step-size is required to better be able to capture the influence the noise, $W_t$, has on the solution, $X_t$. However, this coupling varies with time, so it is not always required to have a small step-size to get a good approximation. Insisting on a small step-size when this is not necessary is a waste of computational resources. An *adaptive* partition exploits this varying coupling, to considerably improve the efficiency of the numerical method used.

In order to know whether the step-size should be made small, one has to know the size of the diffusion coefficient at the next time point. This requires knowledge of the unknown value of the solution at the next time point. To cope with this, one uses a simple explicit method e.g. the Euler-Maruyama method, to compute an approximation to the upcoming value of the solution. This value is then inserted in the main numerical scheme, to correct the step size according to the size of the diffusion coefficient.

This method of predicting a future value, using a simple method, and then inserting this value in a main scheme to make corrections to the structure, is called a *predictor-corrector* method.

Hofmann, Müller-Grönbach and Ritter also prove a theorem that gives a lower bound on the the absolute approximation error, that holds for *any* method that uses a finite number of evaluations of the Wiener process. The adaptive scheme they propose, using the size of the diffusion coefficient as

step size regulator, is shown to obtain this lower bound, thus making it an optimal scheme.

In their article [11], Gaines and Lyons adjusts the step size according to the size of the local error, as the approximate solution is being computed. Their rationale for choosing the step size is that each local error should contribute equally to the global error. They use methods that reduce the mean local error, so that it becomes negligible compared to the standard deviation of the local error. They mention that numerical schemes that are asymptotically efficient have this property.

Since the step size is continuously being adjusted, the Wiener process that drives the stochastic differential equation has to be generated dynamically as the solution is being computed. This contrasts with methods using a fixed step size, where the path of the Wiener process is simulated before the numerical approximation procedure ensues.

Gaines and Lyons use a binary tree structure to be able to store successive refinements of the path of the Wiener process. The need for storage arises from the requirement to be able to compare paths of different resolution. The method of storing a path of the Wiener process in a binary tree, is possible because of the construction of the Wiener process by Lévy [33].

There is interesting unpublished work by Logg, Johnson and Eriksson [34] on ordinary differential equations, in which they propose using small stabilising explicit steps in a numerical method, instead of using an implicit method. They demonstrate the efficiency of their method on a number of stiff problems. It seems that no results of this kind have been published in the context of stochastic differential equations. Numerical methods for stochastic differential equations are often implicit. Hence, this new idea in [34] for ordinary differential equations might, as has happened so often before, generate a useful method in stochastic contexts as well.

## 5.7 Symplectic integration

When one desires a numerical method for approximate solutions of *Hamiltonian* systems, that have high accuracy when integrated over a long period of time, a *symplectic* numerical method is a good choice. A symplectic numerical method replaces the exact Hamiltonian problem with a numerical scheme that gives a solution to an approximate Hamiltonian system.

Symplectic integration has similarities with *geometric integration* which preserves the Lie group structure of a system. See for example the survey article by Channell [7], and the book [51] by Sanz-Serna and Calvo. The connection with stochastic differential equations is through the article by

Tretyakov [43].

Symplectic methods are based on the concepts of wedge product in exterior algebras, and differential forms in complex analysis in several variables. Therefore, we begin by introducing terminology required to work with these concepts.

### 5.7.1 Exterior algebras and differential forms

Here we attempt to give the reader a rudimentary introduction to the topic of exterior algebras, in order to explain symplectic numerical methods. See for example the book by Jacob [25] for a more thorough presentation of exterior algebras.

We begin by defining the concept of a multi-linear functional. This is a generalisation of the concept of a linear functional on a vector space:

**Multi-linear functional** Let $\{V_i\}_{i=1}^n$ be vector spaces over $\mathbb{R}$. A map $f : V_1 \times \ldots \times V_n \to \mathbb{R}$ is called a *multi-linear functional* if

$$\forall i \in \{1, \ldots, n\} : \;\; T_i : V_i \to \mathbb{R},\; T_i(x) = f(x_1, \ldots, x_{i-1}, x, x_i, \ldots, x_n)$$

is a linear functional.

It is a fact that the family, $ML(V_1, \ldots, V_n, \mathbb{R})$, of all such multi-linear functionals is a vector space, with coordinate-wise addition and scalar multiplication.

**Tensor products** If $V$ is a vector space, $V^*$ denotes it dual space, i.e. the space of all bounded linear functionals on $V$. Let $f_i \in V_i^*$. The tensor product $f_1 \otimes \ldots \otimes f_n : V_1 \times \ldots \times V_n \to \mathbb{R}$ is defined by

$$f_1 \otimes \ldots \otimes f_n(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n).$$

In the particular case when the vector spaces $V_i$ are all *finite dimensional*, the tensor product of them is given by

$$\bigotimes^n V = \bigotimes_{j=1}^n V_j = V_1 \otimes \ldots \otimes V_n = ML(V_1^*, \ldots, V_n^*, \mathbb{R}).$$

**Alternating tensor** We say that $f \in \bigotimes^n V^*$ is an *alternating tensor* if

$$f(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) = -f(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n).$$

The family of alternating tensors is denoted $\bigwedge^n V^*$. An element in $\bigwedge^n V^*$ is also called an $n$-form.

Now we come to the important definition of a wedge product. This concept is central in determining whether a numerical method is symplectic or not.

**The wedge product** The wedge product, $f \wedge g$, of $f \in \bigwedge^p V^\star$ and $g \in \bigwedge^q V^\star$, is defined as

$$f \wedge g = \frac{1}{p!\, q!} \sum_{\sigma \in S_{p+q}} \text{sign}(\sigma)\sigma(f \otimes g). \tag{5.17}$$

Here $S_{p+q}$ is the set of all permutations of $\{1, \ldots, p+q\}$, while, for any $f \otimes g \in \bigotimes^{p+q} V^*$,

$$\sigma(f \otimes g)(x_1, \ldots, x_{p+q}) = f \otimes g(x_{\sigma(1)}, \ldots, x_{\sigma(p+q)}).$$

The function $\text{sign}(\sigma)$ is defined as

$$\text{sign}(\sigma) = \begin{cases} 1 & \text{if the number of inversions in } \sigma \text{ is even} \\ -1 & \text{if the number of inversions in } \sigma \text{ is odd} \end{cases}$$

As a simple illustration of the wedge product, consider the following example:

**Example 5.7.1** Pick $f, g \in V^\star$, where $V^\star$ is a finite dimensional vector space. According to the definition (5.17),

$$f \wedge g = \sum_{\sigma \in S_2} \text{sign}(\sigma)\sigma(f \otimes g).$$

Now, in $S_2$ there are only two permutations, $\sigma_1(\{1,2\}) = \{1,2\}$ and $\sigma_2(\{1,2\}) = \{2,1\}$, so that

$$f \wedge g = \text{sign}(\sigma_1)\sigma_1(f \otimes g) + \text{sign}(\sigma_2)\sigma_2(f \otimes g) = \sigma_1(f \otimes g) - \sigma_2(f \otimes g).$$

This gives

$$\begin{aligned} f \wedge g(x_1, x_2) &= \sigma_1(f \otimes g)(x_1, x_2) - \sigma_2(f \otimes g)(x_1, x_2) \\ &= f \otimes g(x_1, x_2) - f \otimes g(x_2, x_1) = f(x_1)g(x_2) - f(x_2)g(x_1) \\ &= \begin{vmatrix} f(x_1) & g(x_1) \\ f(x_2) & g(x_2) \end{vmatrix}. \end{aligned} \tag{5.18}$$

The above example shows that the wedge product is a kind of determinant. This connection can be used to define the *differential 2-form*, used in complex analysis of several variables. The concept of differential 2-form is of vital importance in symplectic integration.

**Differential 2-form** Consider a smooth two dimensional function $(f, g) : \mathbb{R}^2 \to \mathbb{R}^2$.

Associated with this is the *functional determinant*

$$\frac{d(f,g)}{d(x,y)} = \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial y} & \frac{\partial g}{\partial y} \end{vmatrix}.$$

Using the connection between the determinant and the wedge product in (5.18), the differential 2-form $df \wedge dg$ is defined:

$$df \wedge dg(x,y) = \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial g}{\partial x} \\ \frac{\partial f}{\partial y} & \frac{\partial g}{\partial y} \end{vmatrix}.$$

**Symplectic transformation** Consider a transformation

$$\mathbf{f_t} = (f_t^{(1)}, f_t^{(2)}) : \mathbb{R}^2 \to \mathbb{R}^2,$$

that is a smooth function of $t \geq 0$. Let $D_2$ be the open unit disc in $\mathbb{R}^2$, i.e.

$$D_2 = \{x \in \mathbb{R}^2 : |x| < 1\}.$$

The functional determinant

$$\frac{d\mathbf{f_t}}{d\mathbf{x}} = \frac{d(f_t^{(1)}, f_t^{(2)}}{d(x_1, x_2)} = df_t^{(1)} \wedge df_t^{(2)}$$

contains information of the way the area and orientation of $D_2$ changes under $\mathbf{f_t}$. (A negative sign of the determinant means that the orientation is reversed.)

If the transformation $\mathbf{f}$ is such that, for all $t \geq 0$, the oriented area of $D_2$ is unaltered, then $\mathbf{f}$ is said to be a *symplectic transformation.*

This can be expressed using the differential 2-form as

$$\forall t \in [0, \infty) : \ df_t^{(1)} \wedge df_t^{(2)} = df_0^{(1)} \wedge df_0^{(2)}.$$

**Hamiltonian systems** A *Hamiltonian system* of differential equations is given by

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}$$
$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}$$

The function $H = H(\mathbf{p}, \mathbf{q}, \mathbf{t})$, controlling the dynamics of the system, is called the *Hamiltonian* of the system. The coordinates $\mathbf{q} = (\mathbf{q_1}, \ldots, \mathbf{q_d})$ are called *generalised coordinates*, and $\mathbf{p} = (\mathbf{p_1}, \ldots, \mathbf{p_d})$ are their associated *generalised momenta*. Such systems of equations are used to formulate Newton's laws of motion in the framework of analytical mechanics. Hence the name "momenta".

The space $\Xi$ to which $(\mathbf{p}, \mathbf{q})$ belongs is a subset of $\mathbb{R}^{2d}$. It is called the *phase space*.

The following simple example illustrates these concepts:

**Example 5.7.2** A point mass $m$ suspended from a weight-less wire is performing small oscillations around its point of equilibrium. If $q$ is the angular displacement of the straight wire from the vertical, and $p$ is the angular momentum of the point mass, then the Hamiltonian of this system of one degree of freedom, $d$, is

$$H(p, q, t) = T(p) + V(q) = \frac{p^2}{2m} + \frac{kq^2}{2}. \tag{5.19}$$

Since this Hamiltonian does not depend explicitly on time, the corresponding system is autonomous. The Hamiltonian system of differential equations associated with the Hamiltonian (5.19) is given by:

$$\frac{dp}{dt} = -kq$$
$$\frac{dq}{dt} = \frac{p}{m}$$

The solution of this system can be written in matrix form:

$$\begin{pmatrix} p(t) \\ q(t) \end{pmatrix} = \Phi_t \begin{pmatrix} p_0 \\ q_0 \end{pmatrix}.$$

Here the matrix

$$\Phi_t = \begin{pmatrix} \cos \omega t & -m\omega \sin \omega t \\ (m\omega)^{-1} \sin \omega t & \cos \omega t \end{pmatrix}, \quad \omega = \sqrt{\frac{k}{m}}$$

is called the *propagator* of the system (5.19), since it is interpreted as the transformation that propagates the initial configuration $\begin{pmatrix} p(0) \\ q(0) \end{pmatrix}$, to the current configuration $\begin{pmatrix} p(t) \\ q(t) \end{pmatrix}$.

In this example the matrix $\Phi_t$ can be factored

$$\Phi_t = \begin{pmatrix} m\omega & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\omega t & -\sin\omega t \\ \sin\omega t & \cos\omega t \end{pmatrix} \begin{pmatrix} (m\omega)^{-1} & 0 \\ 0 & 1 \end{pmatrix} = A\, R_t\, B.$$

In the phase space, which in this case is $\mathbb{R}^2$, $B$ contracts area by a factor $(m\omega)^{-1}$, while $R_t$ rotates the contracted area an angle $\omega t$, and $A$ expands the area by a factor $m\omega$. The net result is that the transformation induced by the matrix $\Phi_t$ preserves phase space area.

In general, the *propagator* is the transformation $\Phi : (p_0, q_0) \to (p, q)$.

The preservation of phase space area can more conveniently be expressed using differential 2-forms, as is illustrated in the following example:

**Example 5.7.3** Consider the harmonic oscillator

$$\begin{aligned} dp &= P_1\, dp_0 + P_2\, dq_0 = \cos\omega t\, dp_0 - m\omega \sin\omega t\, dq_0 \\ dq &= Q_1\, dp_0 + Q_2\, dq_0 = (m\omega)^{-1} \sin\omega t\, dp_0 + \cos\omega t\, dq_0 \end{aligned}$$

We want to show that the propagator is symplectic:

$$\begin{aligned} dp \wedge dq &= (P_1 Q_2 - P_2 Q_1)\, dp_0 \wedge dq_0 = (\cos^2\omega t + \sin^2\omega t)\, (dp_0 \wedge dq_0) \\ &= dp_0 \wedge dq_0. \end{aligned}$$

This proves the conservation of phase space area. Note that this procedure of proving conservation is more efficient than the factoring of the propagator $\Phi_t$ in the previous example.

It can be shown that the propagator of a Hamiltonian system is symplectic. In fact, the following theorem holds:

**Theorem 5.7.1** *Let the domain $\Xi$ of the operator $\Phi$ be simply connected. Then the only sympletic transformations on $\Xi$ are the propagators of Hamiltonian systems.*

This theorem motivates the use of *symplectic numerical methods*, i.e. numerical methods that preserve the symplectic structure of differential equations.

### 5.7.2 Symplectic numerical methods

Consider solving the Hamiltonian system

$$\begin{aligned} \frac{dp}{dt} &= -\frac{\partial H}{\partial q} \\ \frac{dq}{dt} &= \frac{\partial H}{\partial p} \end{aligned} \tag{5.20}$$

using the numerical method, represented by the operator $\Psi$, given by

$$(p_{n+1}, q_{n+1}) = \Psi_{t_{n+1},t_n}(p_n, q_n).$$

This is to be interpreted as saying that, during the time interval $[t_n, t_{n+1}]$, $\Psi$ evolves the approximate solution $(p_n, q_n)$ to $(p_{n+1}, q_{n+1})$.

The evolution from time $t_n$ to time $t_{n+1}$ of the *exact* solution $(p(t), q(t))$ to the Hamiltonian system, can be represented using the propagator $\Phi$ given by

$$\big(p(t_{n+1}, q(t_{n+1})\big) = \Phi_{t_{n+1},t_n}\big(p(t_n), q(t_n)\big).$$

Notice that $(p_n, q_n)$ is an approximation to $(p(t_n), q(t_n))$. Further, $\Phi$ is symplectic, since it is the propagator of the Hamiltonian system (5.20).

If the operator $\Psi$ is a symplectic transformation, i.e. the numerical scheme is symplectic, and the space $\Xi$ on which the numerical solution exists is simply connected, then Theorem 5.7.1 states that $\Psi$ is the propagator of a Hamiltonian system, whose Hamiltonian function is denoted

$$H_A(\tilde{p}, \tilde{q}, t).$$

This will make $(p_n, q_n) = (\tilde{p}(t_n), \tilde{q}(t_n))$ an exact solution to the Hamiltonian system

$$\begin{aligned} \frac{d\tilde{p}}{dt} &= -\frac{\partial H_A}{\partial \tilde{q}} \\ \frac{d\tilde{q}}{dt} &= \frac{\partial H_A}{\partial \tilde{p}} \end{aligned}$$

This system is an approximation of (5.20).

So what we have when, we have succeeded in designing a symplectic numerical approximation scheme, is an exact solution to a perturbed version of the problem that we were interested in solving in the first place. This idea is called *backward error analysis* in numerical linear algebra. See for example Hairer [14] and Hairer and Lubich [15] for more on the topic of backward error analysis.

### 5.7.3 Symplectic Runge-Kutta methods

A general Runge-Kutta method for numerically solving a system of ordinary differential equations

$$\frac{dy}{dt} = f(y,t)$$

may be represented by the following tableau:

$$\begin{array}{c|ccc} a_{11} & \cdots & a_{1s} \\ \vdots & \ddots & \vdots \\ a_{s1} & \cdots & a_{ss} \\ \hline b_1 & \cdots & b_s \end{array}$$

The entries in this tableau are used to design a numerical method

$$\begin{aligned} y_{n+1} &= \Psi(t_{n+1}, t_n) y_n \\ &= y_n + h_{n+1} \sum_{i=1}^{s} b_i f(\zeta_i, t_n + c_i h_{n+1}), \end{aligned}$$

where

$$\begin{aligned} \zeta_i &= y_n + h_{n+1} \sum_{j=1}^{s} a_{ij} f(\zeta_j, t_n + c_j h_{n+1}) \\ c_i &= \sum_{k=1}^{s} a_{ij} \\ h_{n+1} &= t_{n+1} - t_n \end{aligned}$$

An *explicit* Runge-Kutta method has a lower triangular tableau:

$$\begin{array}{c|cccc} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline . \ b_1 & b_2 & \cdots & b_s \end{array}$$

To design a symplectic Runge-Kutta method, is simply a matter of choosing the right kind of weights, $b_i$ and $a_{ij}$, as the following theorem shows:

**Theorem 5.7.2** *If a Runge-Kutta method has a tableau such that*

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0,$$

*then the Runge-Kutta method is a symplectic method.*

### 5.7.4 Symplectic methods for SDE

To illustrate symplectic methods for stochastic differential equations, consider the following system:

$$\begin{aligned} dP &= f(t,P,Q)\,dt + \sum_{i=1}^{m} \sigma_i(t,P,Q)\,dW_i(t) \\ P(t_0) &= p \\ dQ &= g(t,P,Q)\,dt + \sum_{i=1}^{m} \gamma_i(t,P,Q)\,dW_i(t) \\ Q(t_0) &= q \end{aligned} \tag{5.21}$$

where $\{W_i(t)\}_{i=1}^m$ are independent Wiener processes. If there are functions $H_0(t,p,q)$ and $\{H_i(t,p,q)\}_{i=1}^m$ such that

$$\begin{aligned} f &= -\frac{\partial H_0}{\partial q} \quad g = \frac{\partial H_0}{\partial p} \\ \sigma_i &= -\frac{\partial H_i}{\partial q} \quad \gamma_i = \frac{\partial H_i}{\partial p} \end{aligned}$$

then the phase flow of (5.21) preserves symplectic structure.

For a system where the coefficients $\sigma_i$ and $\gamma_i$ depend only on $t$, and where $f(t,p,q) = f(t,q)$ and $g(t,p,q) = g(p)$, Milstein, Repin and Tretyakov [42] propose an explicit symplectic method, which in the one dimensional case takes the following form:

$$\begin{aligned} P_{k+1} &= P_k + f\big(t_k + h, Q_k + hg(P_k)\big)\,h + \sigma(t_k)\,\Delta_k W \\ Q_{k+1} &= Q_k + g(P_k)\,h + \gamma(t_k)\,\Delta_k W \\ \Delta_k W &= W(t_k + h) - W(t_k) \end{aligned}$$

Here the interval $[0,T]$ is partitioned into equal parts:

$$0 = t_0 < t_1 < \ldots < t_N = T, \quad t_{k+1} - t_k = h. \tag{5.22}$$

This method is of mean-square order 1, i.e. with $\mathbf{X} = \begin{pmatrix} Q \\ P \end{pmatrix}$, and $\mathbf{X}_k$ the

approximation of the exact solution $\mathbf{X}(t_k)$ at $t_k$,

$$\sqrt{\mathbb{E}\big[|\mathbf{X}(t_k) - \mathbf{X}_k|^2\big]} \le Ch.$$

Here is an example that shows some properties of a symplectic method:

$$\begin{aligned} dQ(t) &= P(t)\,dt + \sigma\,dW_1(t) \\ dP(t) &= -Q(t)\,dt + \gamma\,dW_2(t) \end{aligned}$$

With the partition (5.22), the exact solution $\mathbf{X}(t) = \begin{pmatrix} Q(t) \\ P(t) \end{pmatrix}$ is computed as

$$\mathbf{X}(t_{k+1}) = F\mathbf{X}(t_k) + u_k, \tag{5.23}$$

where $\mathbf{X}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $F = \begin{pmatrix} \cos h & \sin h \\ -\sin h & \cos h \end{pmatrix}$, and

$$u_k = \begin{pmatrix} \sigma \displaystyle\int_{t_k}^{t_{k+1}} \cos(t_{k+1} - s)\,dW_1(s) + \gamma \int_{t_k}^{t_{k+1}} \sin(t_{k+1} - s)\,dW_2(s) \\ -\sigma \displaystyle\int_{t_k}^{t_{k+1}} \sin(t_{k+1} - s)\,dW_1(s) + \gamma \int_{t_k}^{t_{k+1}} \cos(t_{k+1} - s)\,dW_2(s) \end{pmatrix}.$$

In [41] Milstein proposes the following symplectic method:

$$\begin{aligned} \mathbf{X}_{k+1} &= H\mathbf{X}_k + v_k \\ H &= \begin{pmatrix} \cos h & \sin h \\ -\sin h & \cos h \end{pmatrix} \\ v_k &= \frac{1}{h} \begin{pmatrix} \sigma \sin h & 2\gamma \sin^2 \frac{1}{2}h \\ -2\sigma \sin^2 \frac{1}{2}h & \sigma \sin h \end{pmatrix} \begin{pmatrix} \Delta W_1 \\ \Delta W_2 \end{pmatrix} \end{aligned} \tag{5.24}$$

The error of this method is of mean-square order 2, i.e.

$$\sqrt{\mathbb{E}\big[|\mathbf{X}(T) - \mathbf{X}_N|^2\big]} \le CTh^2.$$

If there is no noise in the system, then the above symplectic scheme gives the exact solution, since $H = F$. To show the performance of the symplectic method, the following Euler approximation is computed for comparison:

$$\begin{aligned} \mathbf{X}_{k+1} &= \overline{H}\mathbf{X}_k + \overline{v}_k \\ \overline{H} &= \begin{pmatrix} 1 & h \\ -h & 1 \end{pmatrix} \\ \overline{v}_k &= \frac{1}{h} \begin{pmatrix} \sigma & 0 \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} \Delta W_1 \\ \Delta W_2 \end{pmatrix} \end{aligned} \tag{5.25}$$

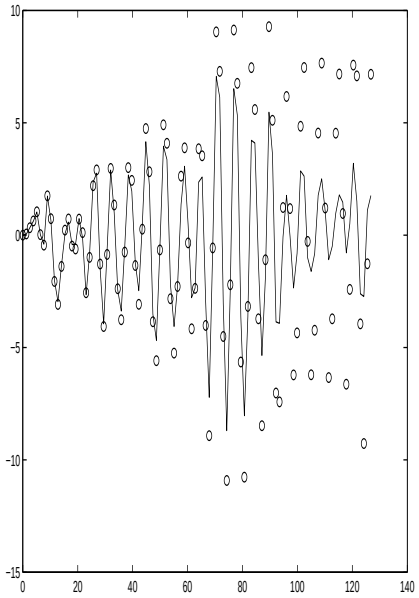


Figure 5.3: The Euler-Maruyama method(∘) (5.25) compared with the exact solution (solid).

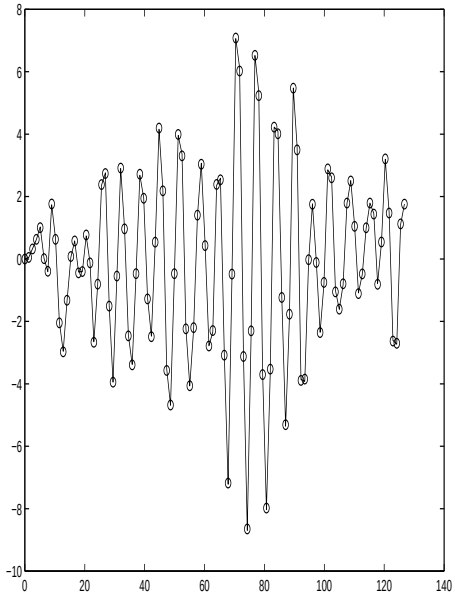


Figure 5.4: The symplectic method (5.24) (∘) compared with the exact solution (solid)

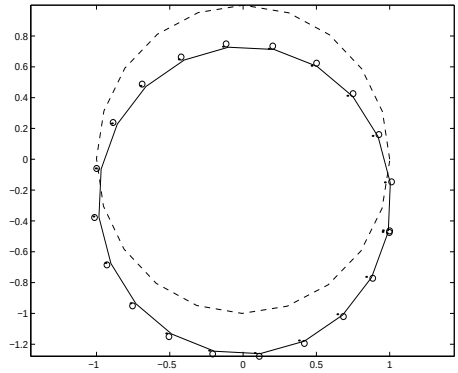


Figure 5.5: Evolution in phase space after 0.2 seconds

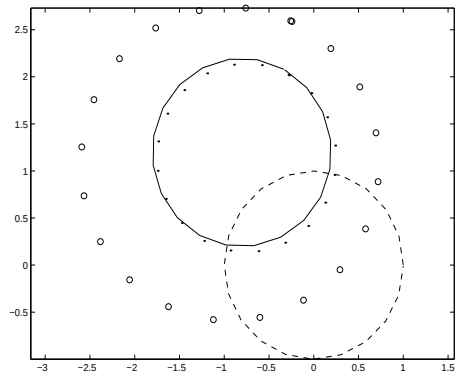


Figure 5.6: Evolution in phase space after 5 seconds

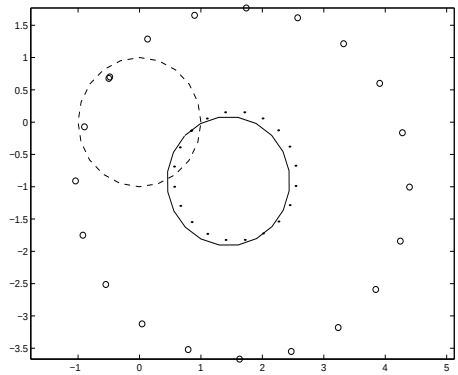


Figure 5.7: Evolution in phase space after 10 seconds

Figure 5.3 shows that the Euler-Maruyama approximation is not good

over a long time interval. In contrast to this, Figure 5.4 shows that the symplectic method (5.24) is very good in approximating the exact solution over a long time interval. Figures 5.5-5.7 show that the symplectic method really is a symplectic method (dotted graph), i.e. it preserves phase space area. The Euler-Maruyama does not preserve phase space area, which is demonstrated by the widening dotted circles. The fact that the Hamiltonian system of equations is stochastic can be seen by noticing that the exact solution (solid circle) is moving around in the phase space. Were the Hamiltonian system to consist of ordinary differential equations, then the exact solution would always move on the circle determined by the initial states. Here the initial configurations lie on the dashed circle.

For the Euler approximation (5.25), the following error estimate holds:

$$\mathbb{E}\big[||\mathbf{X}(T) - \mathbf{X}_N||^2\big] \leq CT^3h^2.$$

## 5.8 Geometric integration

The theory of geometric integration is a part of computational mathematics involved in designing numerical methods for ordinary differential equations evolving on a *differentiable manifold*. A geometric integration scheme aims to make the numerical solution evolve on the same manifold as the exact solution. This can be accomplished if the differentiable manifold is a Lie group. So the theory of geometric integration is about applying the theory of Lie groups to the problem of numerically solving an ordinary differential equation.

The reason for geometric integration appearing in this thesis on numerical methods for stochastic differential equations, is that recently there has been some advances to extend the theory of geometric integration to cover stochastic differential equations as well.

The following example demonstrates that, if the solution of an ordinary differential equation evolves on a manifold, a numerical scheme that respects this, is superior from an accuracy point of view, to a scheme that does not respect this. Also, there is the gained benefit of studying the properties of the numerical solution in order to obtain qualitative information of the exact solution.

**Example 5.8.1** Consider the system of ordinary differential equations

$$\frac{d\mathbf{y}}{dt} = \mathbf{A}\mathbf{y}, \quad \mathbf{y}(\mathbf{0}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{5.26}$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}.$$

The exact solution is

$$\mathbf{y} = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} \sin t \\ \cos t \end{pmatrix},$$

which consequently it lies on the unit circle.

As for a numerical method to solve the system, the following explicit Euler method, obtained simply by replacing the derivative operator by a difference quotient, might be what first comes to mind:

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \mathbf{A}\mathbf{y}_n h = (\mathbf{1} + h\mathbf{A})\mathbf{y}_n.$$

This method gives approximations that do not lie on the unit circle, since the matrix $\mathbf{B} = \mathbf{1} + h\mathbf{A}$ is not orthogonal, and thus does not represent a rotation. However, since $\mathbf{B}\mathbf{B}' = (1 + h^2)\mathbf{1}$, $\mathbf{B}$ is "almost" orthogonal, for $h$ sufficiently small. This is illustrated in Figure 5.8 below.

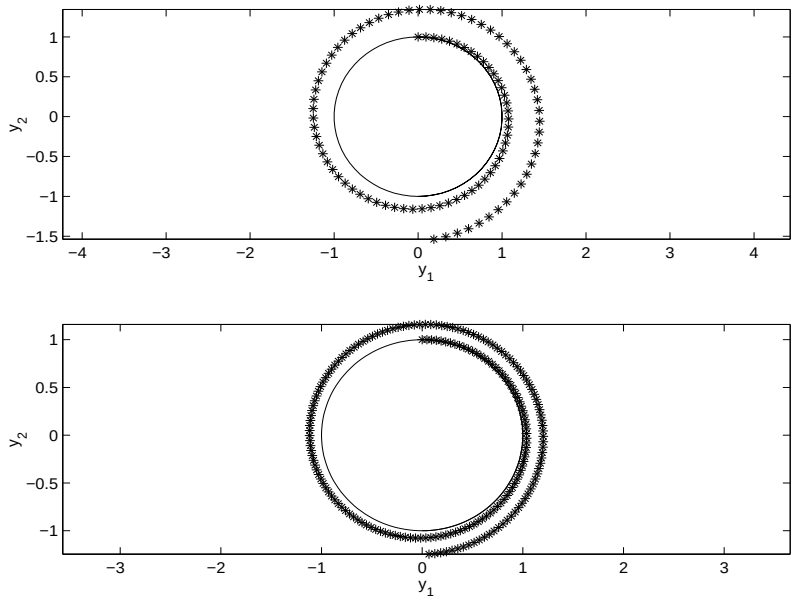



Figure 5.8: The graphs represent the exact solution (solid graph) and the numerical approximation using the explicit Euler method (star-dotted graph) for two different values of the step size $h$ in the phase plane ($y_1$ vs. $y_2$). The top graph has $h = \frac{1}{100}$ and the bottom graph $h = \frac{1}{200}$.

The length of the exact solution vector is invariant under the rotation operator, whereas the length of the Euler approximation vector is *not* invariant under the rotation operator. This suggests the use of a numerical method that makes the length of the solution vector an invariant quantity. An example of such a numerical method is

$$\mathbf{y_{n+1}} = \mathrm{e}^{\mathbf{hA}}\mathbf{y_n}, \tag{5.27}$$

as is shown by Figure 5.9 below.

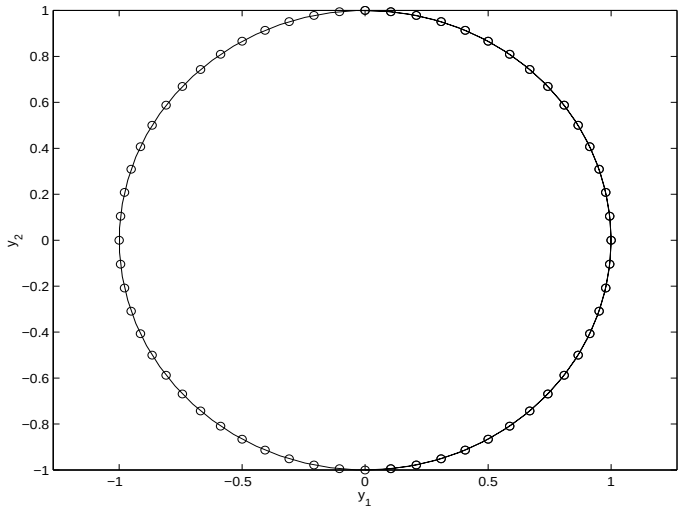



Figure 5.9: The graph shows the exact solution (solid graph) and an approximation, using the numerical method (5.27), in the phase plane ($y_1$ vs. $y_2$). (o-dotted graph).

The unit circle is an example of what is called a differentiable manifold, and the solution vector to the equation (5.26) is seen to evolve on this manifold.

This simple example demonstrates a connection between the structure of a system of differential equations, and an operation under which the solution is invariant.

### 5.8.1 Differential geometry and Lie group theory

In order to be able to understand the subject of geometric integration, some basic background material from differential geometry is required. For more information on geometric integration, see for example the survey by Iserles

et al. in [23]. For more on the theory of Lie groups, see for example the books by Helgason [18], Onishchik and Vinberg [47], and Varadarajan [60].

**Differentiable manifold** Basic to the theory of geometric integration is the concept of a *differentiable manifold*, $\mathcal{M}$. An $n$-dimensional differentiable manifold, $\mathcal{M}$, can be embedded as a surface in $\mathbb{R}^N$, for some $N \geq n$, meaning that when restricted to $\mathcal{M}$, the topological properties of $\mathbb{R}^N$ are the same as the topological properties of $\mathcal{M}$. Thus, a differentiable manifold can be thought of as a surface that locally has a structure like $\mathbb{R}^n$, but globally may differ from such a structure. When talking about manifolds in this thesis, we will always mean differentiable manifolds.

**Tangent to the manifold** Let $\mathcal{M}$ be an $n$-dimensional manifold embedded in $\mathbb{R}^N$, where $N \geq n$. A *tangent* at a point $p \in \mathcal{M}$ is a vector, $v \in \mathbb{R}^N$, such that the distance between $p + \epsilon v$ is of the order $\epsilon^2$, for $\epsilon$ small, i.e. the vector $p + \epsilon v$ lies close to the surface $\mathcal{M}$ if $\epsilon$ is small.

**Tangent vector field** If $p$ is a point on the manifold $\mathcal{M}$, then $T\mathcal{M}_p$ is the set of all tangents at $p$. It is called the *tangent space* at $p$. The set of all tangent spaces of $\mathcal{M}$ is called the *tangent bundle* of $\mathcal{M}$, and is denoted $T\mathcal{M}$. Of course $T\mathcal{M} = \bigcup_p T\mathcal{M}_p$. For a fixed $p$, $T\mathcal{M}$ is a linear space. (This fact will be important when constructing numerical methods for differential equations that *evolve on a manifold*.)

A *tangent vector field* on the manifold $\mathcal{M}$ is a smooth map $F : \mathcal{M} \to T\mathcal{M}$ that associates a tangent vector with a point on $\mathcal{M}$, i.e. $F(p) \in T\mathcal{M}_p$ for all $p \in \mathcal{M}$. The set of all tangent vector fields on $\mathcal{M}$ is denoted $\mathcal{X}(\mathcal{M})$, and is a linear space.

**Differential equations evolving on a manifold** If $F \in \mathcal{X}(\mathcal{M})$ is a tangent vector field on $\mathcal{M}$, a *differential equation evolving on* $\mathcal{M}$ is a differential equation which specifies that the solution $y(t)$ to the differential equation changes along a tangent at $y(t)$, i.e.

$$\frac{dy}{dt} = F(y(t)),\ y(0) \in \mathcal{M}. \tag{5.28}$$

**Flow of a tangent vector field** The solution of a differential equation can be considered to evolve from the initial state $y(0)$ at time $t = 0$, to the state $y(t)$ at time $t$, according to the dynamics specified by the differential equation. The tangent vector field $F$ governs the infinitesimal evolution of $y$. When compounded, these infinitesimal evolutions generate a finite evolution.

Let $\Psi_{t,F}$ denote this finite evolution. It is called the *flow of the tangent vector field* $F$. It is an operator from $\mathcal{M}$ to $\mathcal{M}$. The infinitesimal evolutions of $\Psi_{t,F}$ equals $F$. Thus

$$F = \frac{d}{dt}\Psi_{t,F}.$$

The solution $y(t)$ can be written as

$$y(t) = \Psi_{t,F}(y(0)).$$

Thus, to solve the differential equation (5.28) amounts to find the flow $\Psi_{t,F}$ of $F$. The process of accomplishing this is called the *exponentiation* of the vector field $F$. The notation is

$$\Psi_{t,F} = \exp(tF) = \mathrm{e}^{tF}.$$

The flow is an example of what is called an *exponential map*. The name comes from the following particular case:

**Example 5.8.2** Consider the differential equation

$$\frac{dy}{dt} = Ay(t),\ y \in \mathbb{R}^n,$$

where $A$ is a constant $n \times n$ matrix. The solution is given by $y(t) = \mathrm{e}^{tA}y(0)$. In this example, the manifold is $\mathcal{M} = \mathbb{R}^n$, and the tangent vector field $F = A$, so that $\Psi_{t,F} = \mathrm{e}^{tA}$ corresponds to the matrix exponential.

**Lie group** A Lie group is a manifold $\mathcal{G}$ equipped with an operation $* : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$, satisfying the following properties:

1. $\forall x, y, z \in \mathcal{G}:\ \ x * (y * z) = (x * y) * z$;
2. $\exists e \in \mathcal{G},\ \forall x \in \mathcal{G}:\ \ e * x = x * e = e$;
3. $\forall x \in \mathcal{G},\ \exists x^{-1} \in \mathcal{G}:\ \ x * x^{-1} = x^{-1} * x = e$;
4. $* : (x, y) \to x * y$ and $x \to x^{-1}$ are continuous mappings.

**Lie algebra** A Lie algebra, $\mathfrak{g}$, over a field $K$, where $K$ is $\mathbb{R}$ or $\mathbb{C}$, is a vector space over $K$ on which there is defined a bilinear map $\mathfrak{g} \times \mathfrak{g} \ni (X, Y) \to [X, Y] \in \mathfrak{g}$, called the *commutator*, having the following properties:

1. $\forall X, Y, Z \in \mathfrak{g}:\ \ [X, [Y, Z]] + [Y, [X, Z]] + [Z, [X, Y]] = 0$;

2. $\forall X \in \mathfrak{g}: \ [X, X] = 0.$

The first property is called the *Jacobi identity.*

**Example 5.8.3** A simple example of a Lie algebra on $\mathbb{R}$, is the set of all vectors in $\mathbb{R}^3$, with the bilinear map between vectors chosen as the vector cross-product, $u \times v$.

**Lie algebra generated by tangent vector fields** A Lie algebra, over $\mathbb{R}$, of tangent vector fields, is a family of tangent vector fields $F = \{F_i\}_{i \in I}$ that is a linear space and closed under commutation, i.e.

$$\forall a, b \in \mathbb{R}, \ \forall F_i, F_j \in F: \ aF_1 + bF_2 \in F;$$

$$\forall F_i, F_j \in F: \ [F_i, F_j] \in F.$$

*The Lie algebra, $\mathfrak{g}$, generated by the tangent vector fields $\{F_i\}_{i \in I}$,* is the smallest Lie algebra of tangent vector fields containing $F$.

The importance of Lie algebras, $\mathfrak{g}$, generated by a family, $F$, of tangent vector fields, lies in the fact that *the flow of any tangent vector field in $\mathfrak{g}$ can be approximated arbitrarily well by composing flows of tangent vector fields in $F$.*

**Finite dimensional Lie algebra** If $\mathfrak{g}$ is a Lie algebra generated by the family of tangent vector fields $\{F_i\}_{i \in I}$, and $I$ is a finite index set, $I = \{1, 2, \ldots, n\}$, then $\mathfrak{g}$ is of finite dimension $n$.

**The Lie algebra of a Lie group** The Lie algebra, $\mathfrak{g}$, of a Lie group $\mathcal{G}$, is the linear space consisting of all tangents of $\mathcal{G}$ at the identity element, $e$, of $\mathcal{G}$. The bilinear map on the Lie algebra $\mathfrak{g}$ is defined by

$$[X, Y] = \frac{\partial^2}{\partial s \partial t} \rho(s)\sigma(t)\rho(-s)\Big|_{s=t=0}$$

for $X, Y \in \mathfrak{g}$, where $\rho(\cdot)$ and $\sigma(\cdot)$ are parametrisations of two smooth curves on $\mathcal{G}$ such that

$$\begin{aligned} \rho(0) &= e = \sigma(0) \\ \frac{d\rho}{ds}\Big|_{s=0} &= X \\ \frac{d\sigma}{dt}\Big|_{t=0} &= Y \end{aligned}$$

**Matrix Lie group** A *real matrix Lie group* is a smooth subset of $\mathbb{R}_{n|n}$, the set of all real $n \times n$ matrices, that is closed under matrix products and matrix inversion.

**Matrix Lie algebra** The Lie algebra, $\mathfrak{g}$, of a matrix Lie group $\mathcal{G}$, is the linear subspace of $\mathbb{R}_{n|n}$, consisting of all matrices of the form

$$A = \frac{d\rho(s)}{ds}\Big|_{s=0},$$

where $\rho(\cdot) \in \mathcal{G}$ is a smooth curve such that $\rho(0) = \mathbf{1}$, with $\mathbf{1}$ the unit matrix in $\mathbb{R}_{n|n}$. The Lie algebra $\mathfrak{g}$ is closed under matrix addition, scalar multiplication, and under the matrix commutator

$$[A, B] = AB - BA.$$

The following quote is from from Iserles et al. in [23]:

"For practically any concept in general Lie theory, there exists a corresponding concept within matrix Lie theory. Vice versa, practically any result that holds in the matrix case remains valid within the general Lie theory."

The truth of this is underscored by the following theorem of Ado (see for example Varadarajan's book [60] on a proof):

**Theorem 5.8.1** *Every Lie algebra of dimension $n$ is isomorphic to a subalgebra of the matrix algebra* $\mathrm{gl}(n)$ *consisting of all $n \times n$ real matrices.*

The observations above are crucial when it comes to the numerical solution of an ordinary differential equation that evolves on a Lie group. Since any numerical solution is based on a discretization, the possibly infinite dimensional Lie group will have to be approximated by a finite dimensional Lie group, to which is associated a finite dimensional Lie algebra.

**Differential equation evolving on a matrix Lie group** If $\mathcal{G}$ is a matrix Lie group, and $\mathfrak{g}$ an associated matrix Lie algebra, then a differential equation evolving on $\mathcal{G}$ is an equation of the form

$$\begin{aligned} \frac{dY(t)}{dt} &= A(t, Y(t))Y(t) \\ Y(0) &= Y_0 \in \mathcal{G} \\ t \in [0, T] & \end{aligned} \tag{5.29}$$

Here $A$ is a map $A : \mathbb{R} \times \mathcal{G} \to \mathfrak{g}$, and $A(t, Y(t))Y(t)$ is the matrix $A(t, Y(t)) \in \mathfrak{g}$ multiplied with $Y(t) \in \mathcal{G}$.

For small $t \geq 0$, the solution of (5.29) is given by

$$Y(t) = \exp(\Theta(t))Y_0,$$

where $\Theta(t) \in \mathfrak{g}$ satisfies the differential equation

$$\begin{aligned} \frac{d\Theta(t)}{dt} &= \operatorname{dexp}^{-1}_{\Theta(t)}(A(t, Y(t))) \\ \Theta(0) &= \mathbf{0} \end{aligned} \tag{5.30}$$

and $\mathbf{0}$ is the zero matrix in $\mathbb{R}_{n|n}$. The operators $\exp(\cdot)$, $\operatorname{dexp}_{(\cdot)}(\cdot)$ and $\operatorname{dexp}^{-1}_{(\cdot)}(\cdot)$ are defined below.

The equation (5.30) was originally stated by Felix Hausdorff in 1906 [17].

**Definition 5.8.1** *The* exponential map, $\exp : \mathfrak{g} \to \mathcal{G}$, *is defined by*

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

**Definition 5.8.2** *The* adjoint *representation,* $\operatorname{Ad}_B(A)$, *and its derivative,* $\operatorname{ad}_B(A)$, *are defined by*

$$\begin{aligned} \operatorname{Ad}_B(A) &= BAB^{-1} \\ \operatorname{ad}_B(A) &= AB - BA = [A, B] \end{aligned}$$

**Definition 5.8.3** *The differential,* $\operatorname{dexp} : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, *of the exponential map* $\exp : \mathfrak{g} \to \mathcal{G}$, *is defined by the equation*

$$\frac{d \exp(A(t))}{dt} = \operatorname{dexp}_{A(t)}\left(\frac{dA(t)}{dt}\right) \exp(A(t)).$$

The operator $\operatorname{dexp}_A(\cdot)$ is an analytic function of $\operatorname{ad}_A$, that is given by

$$\operatorname{dexp}_A(\cdot) = \frac{\exp(\operatorname{ad}_A(\cdot)) - \mathbf{1}}{\operatorname{ad}_A(\cdot)}.$$

Thus $\operatorname{dexp}_A(\cdot)$ can be inverted, to get a formal inverse $\operatorname{dexp}^{-1}_A(\cdot)$:

$$\operatorname{dexp}^{-1}_A(\cdot) = \frac{\operatorname{ad}_A(\cdot)}{\exp(\operatorname{ad}_A(\cdot)) - \mathbf{1}}.$$

By expanding the function $f(x) = x/(\mathrm{e}^x - 1)$ in a Taylor series, we obtain

$$\frac{x}{\mathrm{e}^x - 1} = \sum_{k=0}^{\infty} \frac{B_k}{k!} x^k.$$

Here $B_k = f'(x)|_{x=0}$ are the so called Bernoulli numbers. More specifically, we have

$$\mathrm{dexp}_A^{-1}(B) = B - \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] + \cdots = \sum_{k=0}^{\infty} \frac{B_k}{k!} \mathrm{ad}_A^{(k)}(B), \quad (5.31)$$

where

$$\mathrm{Ad}_A^{(k)}(B) = \underbrace{[A,[A,\ldots,[A}_{\text{k times}},B]\ldots]],$$

i.e.

$$\begin{aligned} \mathrm{Ad}_A^0(B) &= B \\ \mathrm{Ad}_A^1(B) &= [A,B] \\ \mathrm{Ad}_A^2(B) &= [A,[A,B]] \\ &\vdots \end{aligned}$$

### 5.8.2 The Runge-Kutta-Munthe-Kaas scheme

A numerical solution of (5.30) can be attempted using a *Runge-Kutta* type scheme.

An $m$-stage Runge-Kutta scheme, used to obtain a numerical approximation to a differential equation

$$\begin{aligned} \mathbf{y}' &= \mathbf{f}(t,\mathbf{y}) \\ \mathbf{y}(0) &= \mathbf{y}_0,\ t \in [0,T] \end{aligned}$$

can be represented by a so called *Butcher tableau* [6]:

$$\begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ c_m & a_{m1} & \cdots & a_{mm} \\ \hline & b_1 & \cdots & b_m \end{array}$$

This tableau is shorthand notation for the following numerical scheme:

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \sum_{k=0}^{m} b_k \mathbf{f}_k$$

$$\begin{aligned} \mathbf{f}_k &= h\mathbf{f}(t_n + c_k h, \theta_k) \\ \theta_k &= \mathbf{y}_n + \sum_{i=0}^{m} a_{ki}\mathbf{f}_i \\ h &= t_{n+1} - t_n \\ 0 = t_0 &< t_1 < \ldots < t_N = T \end{aligned} \tag{5.32}$$

When (5.32) is applied to the equation (5.30), the resulting numerical scheme is called the Runge-Kutta-Munthe-Kaas (RKMK) scheme:

$$\begin{aligned} Y_{n+1} &= \exp(\Theta) Y_n \\ \Theta &= \sum_{k=0}^{m} b_k F_k \\ F_k &= \operatorname{dexp}_{\Theta_k}^{-1}(A_k) \\ \Theta_k &= \sum_{i=0}^{m} a_{ki} F_i \\ A_k &= hA(t_n + c_k h, \exp(\Theta_k) Y_n) \\ h &= t_{n+1} - t_n \\ 0 = t_0 &< t_1 < \ldots < t_N = T \end{aligned} \tag{5.33}$$

In order to apply this scheme, it is necessary to have an approximation for the operator $\operatorname{dexp}^{-1}$. One way to approximate $\operatorname{dexp}^{-1}$ is to truncate the expansion (5.31). If the RKMK scheme is to be of the same order as the underlying Runge-Kutta scheme, the truncation has to be of the same order as that of the underlying Runge-Kutta scheme. For high-order methods, this implies the use of several commutators from (5.31), which gives very complex numerical methods, see [8]. One way of reducing the required number of commutators is to use so called *free Lie algebras*. For more on this, see the survey article on Lie group methods by Iserles et al. [23].

An important fact is that, even if $\operatorname{dexp}^{-1}$ is approximated, the resulting numerical approximation, obtained by (5.33), will remain in the same Lie group as the exact solution. This is the main idea with geometric integration methods.

### 5.8.3 The Magnus formula

Previously we saw that the problem of solving the equation

$$Y' = AY$$

$$Y_0 \in \mathcal{G} \quad \left(Y' = \frac{dY}{dt}\right)$$

was reduced to solve the associated equation $\Theta' = \mathrm{dexp}_{\Theta}^{-1}(A)$ for the exponent, in the representation $Y = \exp(\Theta)Y_0$.

One way to approach the solution of this equation is by way of *Picard iteration*:

$$\Theta_0(t) = \mathbf{0}$$
$$\Theta_n(t) = \int_0^t \mathrm{dexp}_{\Theta_{n-1}(s)}^{-1}(A(s))\, ds$$

In a suitably small neighbourhood of $\mathbf{0}$, this scheme converges to $\Theta(t) = \lim_{n\to\infty} \Theta_n(t)$. The solution $\Theta(t)$ can be represented with the *Magnus expansion* or the *Magnus formula*. This is a linear combination of iterated integrals and commutators:

$$\Theta(t) = \sum_{k=0}^{\infty} H_k(t), \tag{5.34}$$

where

$$\begin{aligned}
H_0(t) &= \int_0^t A(s_1)\, ds_1 \\
H_1(t) &= -\frac{1}{2}\int_0^t \left[\int_0^{s_1} A(s_2)\, ds_2, A(s_1)\right] ds_1 \\
H_2(t) &= \frac{1}{12}\int_0^t \left[\int_0^{s_1} A(s_2)\, ds_2, \left[\int_0^{s_1} A(s_2)\, ds_2, A(s_1)\right]\right] ds_1 \\
&\quad + \frac{1}{4}\int_0^t \left[\int_0^{s_1}\left[\int_0^{s_2} A(s_3)\, ds_3, A(s_2)\right] ds_2, A(s_1)\right] ds_1 \\
&\ \ \vdots
\end{aligned} \tag{5.35}$$

As can be seen, the terms in the Magnus expansion become increasingly more complex. A more transparent form of the expansion can be obtained by associating each term with a *rooted binary tree*. This approach makes numerical computations and recursive derivations of additional terms in the expansion more tractable. Thus, the use of *graph theory* becomes involved.

We will not go into this topic any further here, except to point the interested reader to some literature: the article by Iserles and Norsett [24], the book by Butcher [6], which addresses the method of rooted trees, and the book by Harary [16], on general graph theory.

It was mentioned that the Picard iteration scheme, that generated the

Magnus expansion, holds in a suitably small neighbourhood of zero. [Recall that $\Theta(0) = \mathbf{0}$.] The following theorem by Moan [45] specifies exactly how small this neighbourhood is:

**Theorem 5.8.2** *If the Lie algebra $g$ is equipped with a norm, $||\cdot||$, the Magnus expansion is absolutely convergent with respect to this norm, for every $t \geq 0$ that satisfies*

$$\int_0^t ||A(s)||\, ds \leq \int_0^{2\pi} \frac{ds}{4 + s(1 - \cot \frac{s}{2})} \approx 1.089.$$

The Magnus expansion is, of course, used to obtain approximative solutions to (5.30) by truncation. The truncation cannot be performed just anywhere in the expansion. The rooted tree approach gives information on where to truncate, and on the order of the approximation thus obtained. Here we give a few truncations that can be used to construct numerical schemes:

$$\Theta(t) \approx \int_0^t A(s_1)\, ds_1 \qquad \text{Order 2}$$

$$\Theta(t) \approx \int_0^t A(s_1)\, ds_1 - \frac{1}{2}\int_0^t \left[\int_0^{s_1} A(s_2)\, ds_2, A(s_1)\right] ds_1 \qquad \text{Order 4}$$

Once these truncations have been made, the integrals are approximated by so called numerical *quadrature*. This is as far as we will go on this subject here, since the aim merely is to give the reader some *ideas* of the methods involved in computing numerical solutions of ordinary differential equations, that evolve on differentiable manifolds. For more on numerical quadrature, consult for example Shampine [53], or any other book on numerical solution of ordinary differential equations.

### 5.8.4 Geometric integration and SDE: The Magnus formula

In [5], Burrage and Burrage apply the Magnus formula to an SDE in Stratonvich form. This seems to be the first paper in which the Magnus formula is used in a stochastic context.

Burrage and Burrage consider the following vector valued SDE:

$$\begin{aligned} dX(t) &= G_0 X(t)\, dt + G_1 X(t) \circ dW_1(t) + G_2 X(t) \circ dW_2(t) \\ X(0) &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{aligned} \tag{5.36}$$

$$t \in [0, 1]$$

Here $W_1$ and $W_2$ are independent Wiener processes, while $G_0$, $G_1$ and $G_2$ are the $2 \times 2$-matrices, given by

$$G_0 = \begin{pmatrix} -\frac{9}{10} & 0 \\ \frac{1}{4} & -\frac{1}{2} \end{pmatrix}, \quad G_1 = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & -\frac{3}{4} \end{pmatrix}, \quad \text{and} \quad G_2 = \begin{pmatrix} 0 & \frac{9}{10} \\ \frac{9}{10} & 0 \end{pmatrix}.$$

The matrices $G_0$, $G_1$ and $G_2$ do not commute, so the Magnus expansion is needed to obtain a good numerical approximation.

The solution to (5.36) is

$$X(t) = \exp(\Theta(t))X(0),$$

with $\Theta$ given by (5.34) and (5.35). In our example (5.36)

$$A(t)\,dt = G_0\,dt + G_1 \circ dW_1(t) + G_2 \circ dW_2(t).$$

Thus the first term in the Magnus expansion (5.35) is

$$H_0(t) = \int_0^t A(s)\,ds = G_0 t + G_1 J_1 + G_2 J_2,$$

where

$$J_1 = \int_0^t dW_1(s) = W_1(t) \quad \text{and} \quad J_2 = \int_0^t dW_2(s) = W_2(t).$$

The second term can be computed, with somewhat more effort,

$$H_1(t) = [G_0, G_1](J_{10} - J_{01}) + [G_0, G_2](J_{20} - J_{02}) + [G_1, G_2](J_{21} - J_{12}),$$

where

$$J_{ij} = \int_0^t \int_0^{s_1} dW_i(s_2) \circ dW_j(s_1)$$

and

$$[G_i, G_j] = G_i G_j - G_j G_i.$$

In order to obtain numerical solutions to the SDE (5.36), it is necessary to simulate the iterated Stratonovich integrals $J_{ij}$ above. One way of doing this is described in the article by Ryden and Wiktorsson [50].

Notice that the iterated integrals appearing in (5.35) are interrelated. See [29] for an extensive discussion of the interrelationships.

Simulations of iterated integrals of higher order, when more terms are needed, are more involved, of course. Using the mentioned interrelations,

these simulation can be made more tractable, as shown by Milstein [41].

Using only $H_0$ and $H_1$, Burrage and Burrage [5] construct a numerical scheme based on the Magnus expansion, and compare it to the Euler scheme. The results are depicted in Figure 5.10 and Figure 5.11.

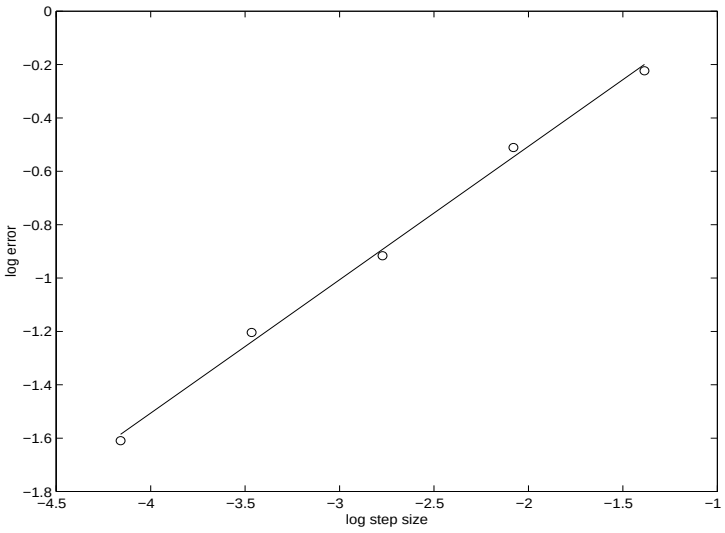



Figure 5.10: The logarithm of the global error versus the logarithm of the step size. The error using the Euler scheme (o) follows a line with slope 0.5, indicating that the scheme is of strong global order 0.5.

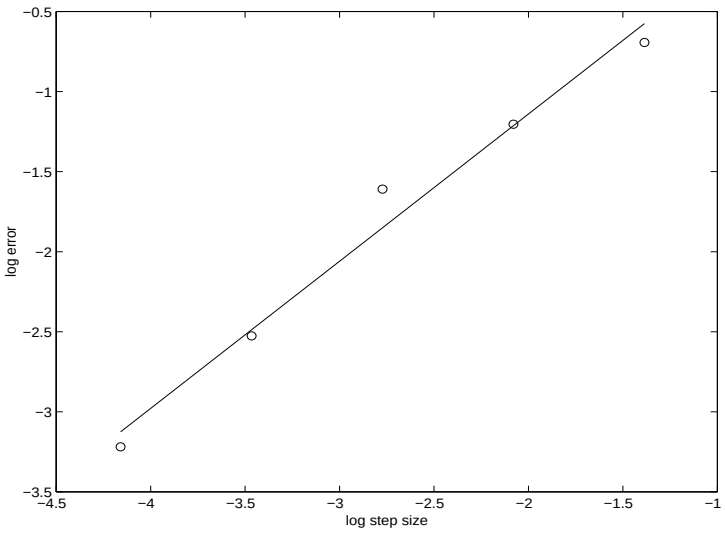



Figure 5.11: The logarithm of the global error versus the logarithm of the step size. The error using the scheme based on the Magnus expansion (o) follows a line with slope 1.0, indicating that the scheme is of strong global order 1.0.

### 5.8.5 Geometric integration and SDE: The Kunita formula

Consider the d-dimensional SDE in Stratonovich form, driven by a one dimensional Wiener-process $W(t)$:

$$\begin{aligned} dS(t) &= b(S(t))\,dt + g(S(t)) \circ dW_t \\ S(0) &= s \end{aligned} \tag{5.37}$$

Here the solution process $S(t)$ is assumed to evolve in a $d$-dimensional connected $C^\infty$-manifold, $\mathcal{M}$,

$$\begin{aligned} S &: [0,\infty) \to \mathcal{M} \\ b\,, g &\in C^\infty(\mathcal{M}, \mathbb{R}^d) \end{aligned}$$

Introduce the operators $X_0$ and $X_1$:

$$X_0 = \sum_{i=1}^{d} b_i \frac{\partial}{\partial s_i} \quad \text{and} \quad X_1 = \sum_{i=1}^{d} g_i \frac{\partial}{\partial s_i}.$$

The following formula by Kunita gives the solution of the SDE (5.37):

**Theorem 5.8.3** *Suppose that the Lie algebra $L(X_0, X_1)$ generated by the operators $X_0$ and $X_1$ is p-nilpotent. Then the solution of* (5.37) *is*

$$S(t) = \mathrm{e}^{Y_t}(S(0)) = \mathrm{e}^{Y_t}(s),$$

*where the operator $Y_t$ is given by the following formula:*

$$\forall t \geq 0: \quad \mathbb{P}\left( Y_t = \sum_{i=0}^{1} W_i(t) X_i + \sum_{J:2\leq|J|\leq p} \sum_{\Delta J}^{*} c_{\Delta J} W_{\Delta J}(t) X_J \right) = 1,$$

*where*

$$\begin{aligned} W_0(t) = t,\ W_1(t) &= W_t \\ J = (j_1, \ldots, j_m),\ |J| &= j_1 + \ldots + j_m \\ X_J &= [[\ldots[X_{j_1}, X_{j_2}]\ldots]X_{j_m}] \\ W_{\Delta J} &= \int \ldots \int dW_{J_1}(t_1) \circ \ldots \circ dW_{J_{k_l}}(t_l) \\ W_{J_k} &= \begin{cases} W_{i_k} & J_k = \{i_k\} \\ t & J_k = \{i_k, i_k\} \\ 0 & J_k = \{0,0\} \end{cases} \end{aligned}$$

The multi-index $J$ is divided into shorter multi-indices, $J = (J_1, \ldots, J_q)$ such that each $J_k$ contains the same element. A further sub-division, $\Delta J = (J_1, \ldots, J_{k_1}, J_{k_1+1}, \ldots, J_{k_2}, \ldots, J_{k_{l-1}+1}, \ldots, J_{k_l})$ is such that each $J_{k_i}$ contain at most two elements, where $k_1 < k_2 < \ldots < k_l = q$. $\Delta J$ is called a single divided multi-index of $J$ if all the $J_{k_i}$ contain a single element. Otherwise $\Delta J$ is called a double divided multi-index of $J$.

The sum $\sum^*_{\Delta J}$ denotes summation over all single and double divided multi-indices of $J$.

The coefficients $c_{\Delta J}$ in the theorem are computed from quite a complicated formula that is not given here. For a description, see the article by Misawa [44].

If the Lie algebra $L(X_0, X_1)$ is *not* $p$-nilpotent, then the operator $Y_t$ is given by an infinite series:

$$Y_t = \sum_{i=0}^{1} W_i(t) X_i + \sum_{J:2\le|J|} \sum_{\Delta J}^{*} c_{\Delta J} W_{\Delta J}(t) X_J, a.s.$$

(Note the absence of $p$ in the second summation.)

If $L(X_0, X_1)$ is of finite dimension, then Arous [1] has proven that there is an $\mathcal{F}$-stopping time $T^*$, where $\mathcal{F}$ is the filtration $\{\mathcal{F}_s\}_{s\ge 0}, \mathcal{F}_s = \sigma(S(r) : 0 \le r \le s)$, such that

$$\forall t \le T^* : \quad \mathbb{P}\left( Y_t = \sum_{i=0}^{1} W_i(t) X_i + \sum_{J:2\le|J|} \sum_{\Delta J}^{*} c_{\Delta J} W_{\Delta J}(t) X_J \right) = 1$$

Kunita's formula implies that the solution $S(t) = \mathrm{e}^{Y_t}(s)$ can be expressed $\mathrm{e}^{Y_t}(s) = \phi(1, s)$, where $\phi(\tau, s)$ satisfies the following ordinary differential equation:

$$\begin{aligned} \frac{d\phi}{d\tau} &= Y_t(\phi) \\ \phi(0, s) &= s \end{aligned} \tag{5.38}$$

### 5.8.6 Applying the Kunita formula

An approximate solution over the interval $[0, T]$ is sought. Discretize the interval into $N$ equal parts of length $\Delta t = T/N$. Let $t_n = n\Delta t$ denote the discretization points, and $S_n = S(t_n)$ the value of the exact solution at these points. Further, let $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ be the increments of the driving Wiener process. The numerical approximation is

$$S_{n+1} = \mathrm{e}^{Y_{n\Delta t}}(S_n),$$

where

$$
\begin{aligned}
Y_{n\Delta t} &= I_{(0),n}X_0 + I_{(1),n}X_1 + \frac{1}{2}(I_{(0,1),n} - I_{(1,0),n})[X_0, X_1] + \dots \\
I_{(0),n} &= \int_{\tau_1=(n-1)\Delta t}^{n\Delta t} dY_0(\tau_1) = \Delta t \\
I_{(1),n} &= \int_{\tau_1=(n-1)\Delta t}^{n\Delta t} dY_1(\tau_1) = \Delta W_n \\
I_{(0,1),n} &= \int_{\tau_2=(n-1)\Delta t}^{n\Delta t} \int_{\tau_1=(n-1)\Delta t}^{\tau_2} dY_0(\tau_1) dY_1(\tau_2) \\
I_{(1,0),n} &= \int_{\tau_2=(n-1)\Delta t}^{n\Delta t} \int_{\tau_1=(n-1)\Delta t}^{\tau_2} dY_1(\tau_1) dY_0(\tau_2) \\
dY_i(t) &= \begin{cases} dt & i = 0 \\ dW_t & i = 1 \end{cases}
\end{aligned}
$$

The infinite series is truncated to obtain $\hat{Y}_{n\Delta t}$. An example of a truncation is

$$\hat{Y}_{n\Delta t} = I_{(0),n}X_0 + I_{(1),n}X_1.$$

In order to compute the exponential of $\hat{Y}_{n\Delta t}$, it is split into $A = A_{n\Delta t}$ and $B = B_{n\Delta t}$, where the exponentials of $A$ and $B$ can be computed explicitly using (5.38). Thus

$$\mathrm{e}^{\hat{Y}_{n\Delta t}} = \mathrm{e}^{A+B}.$$

The Baker-Campbell-Hausdorff formula states the following relation:

$$\mathrm{e}^A \mathrm{e}^B = \exp\left\{A + B + \frac{1}{2}[A, B] + \frac{1}{12}([A, [A, B]] + [B[B, A]] + \dots)\right\}.$$

The exponential $\mathrm{e}^{A+B}$ can be approximated by $\mathrm{e}^A \mathrm{e}^B$ according to this formula. The approximation becomes exact if the operators $A$ and $B$ commute, i.e. $[A, B] = 0$.

In conclusion, the numerical scheme for approximating the solution $S(t)$ reads:

$$\tilde{S}_{n+1} = \mathrm{e}^{A_{n\Delta t}} \mathrm{e}^{B_{n\Delta t}}(\tilde{S}_n),$$

where $\tilde{S}_n$ is an approximation of $S_n = S(t_n)$.

Here is an example that illustrates the use of Kunita's formula:

**Example 5.8.4** Consider the Bessel-type SDE

$$\begin{aligned} dS(t) &= S(t)\,dt + 2\sqrt{S(t)} \circ dW_t \\ S(0) &= s \end{aligned} \tag{5.39}$$

This equation is sometimes used in mathematical finance to model the price of a stock, see for example the work by Geman and Yor [13].

The operators $X_0$ and $X_1$ are

$$\begin{aligned} X_0 &= s\frac{d}{ds} \\ X_1 &= 2\sqrt{s}\frac{d}{ds} \end{aligned}$$

The Lie algebra $L(X_0, X_1)$ generated by $X_0$ and $X_1$ is finite dimensional since $[X_0, X_1] = -\frac{1}{2}X_1$, as demonstrated by the following calculation:

$$[X_0, X_1] = X_0X_1 - X_1X_0 = s\frac{d}{ds}(2\sqrt{s}\frac{d}{ds}) - 2\sqrt{s}\frac{d}{ds}(s\frac{d}{ds}) = -\sqrt{s}\frac{d}{ds} = -\frac{X_1}{2}.$$

The truncated Kunita formula is

$$Y_{n\Delta t} = \Delta t X_0 + \Delta W_n X_1.$$

Let the operators $A_{n\Delta t}$ and $B_{n\Delta t}$ be given by

$$A_{n\Delta t} = \Delta t X_0 \quad \text{and} \quad B_{n\Delta t} = \Delta W_n X_1.$$

Then the exponential map $\mathrm{e}^{A_{n\Delta t}}$ can be computed in the following way:

$$Z_A(t) = \mathrm{e}^{A_{n\Delta t}}(z_A(0)) = \phi_A(1),$$

where the function $\phi_A(\tau)$ satisfies the ordinary differential equation

$$\begin{aligned} \frac{d\phi_A(\tau)}{d\tau} &= A_{n\Delta t}(\phi_A(\tau)) = \Delta t \phi_A \frac{d\phi_A}{d\phi_A} \\ \phi_A(0) &= z_A(0) \end{aligned}$$

The solution is $\phi_A(\tau) = z_A(0)\mathrm{e}^{\tau\Delta t}$, which gives $\mathrm{e}^{A_{n\Delta t}}(s) = s\mathrm{e}^{\Delta t}$.

As for $\mathrm{e}^{B_{n\Delta t}}$, this quantity can be computed in the following way:

$$Z_B(t) = \mathrm{e}^{B_{n\Delta t}}(z_B(0)) = \phi_B(1),$$

where the function $\phi_B(\tau)$ satisfies the ordinary differential equation

$$\frac{d\phi_B(\tau)}{d\tau} = B_{n\Delta t}(\phi_B(\tau)) = \Delta W_n 2\sqrt{\phi_B}\,\frac{d\phi_B}{d\phi_B}$$
$$\phi_B(0) = z_B(0)$$

This gives $\phi_B(\tau) = (\tau\Delta W_n + \sqrt{z_B(0)})^2$, so that $\mathrm{e}^{B_{n\Delta t}}(s) = (\Delta W_n + \sqrt{s})^2$. In conclusion, we have the following numerical approximation scheme:

$$\tilde{S}_{n+1} = \left(\Delta W_n + \sqrt{\tilde{S}_n}\right)^2 \mathrm{e}^{\Delta t}. \tag{5.40}$$

In [44], Misawa shows that this scheme is of mean-square order 1.

The solution, $S(t)$, to the SDE (5.39) can be shown to be non-negative. The numerical scheme derived above shares this property. This is not the case for e.g., the Euler-Maruyama scheme,

$$\bar{S}_{n+1} = \bar{S}_n\, h + 2\sqrt{\bar{S}_n}\,\Delta W_n,$$

which is of mean-square order 0.5.

Since the exact solution of (5.39) cannot be obtained in closed form, in order to compare the numerical solution based on the Kunita formula with that obtained by the Euler-Maruyama method, a high order approximation, based on a stochastic Taylor formula, is used as a substitute for the exact solution. This high order approximation is a method of mean-square order 1.5, suggested by Kloeden [29]:

$$\begin{aligned} S_{n+1} = {} & S_n + (S_n + 1)\, h + 2\sqrt{S_n}\,\Delta W_n \\ & + (\Delta^2 W_n - h) + 2\sqrt{S_n}\, I_{(1,0),n}(h) + \sqrt{S_n}\, I_{(0,1),n}(h) \\ & + \frac{1}{2}(S_n + 1)\, h^2. \end{aligned} \tag{5.41}$$

Here

$$I_{(1,0),n}(h) = \frac{1}{2}\left(\xi_n + \frac{1}{\sqrt{3}}\zeta_n\right) h^{3/2}$$

and

$$I_{(0,1),n}(h) = \frac{1}{2}\left(\xi_n - \frac{1}{\sqrt{3}}\zeta_n\right) h^{3/2},$$

where $\xi_n$ and $\zeta_n$ are independent $N(0,1)$-distributed random variables.

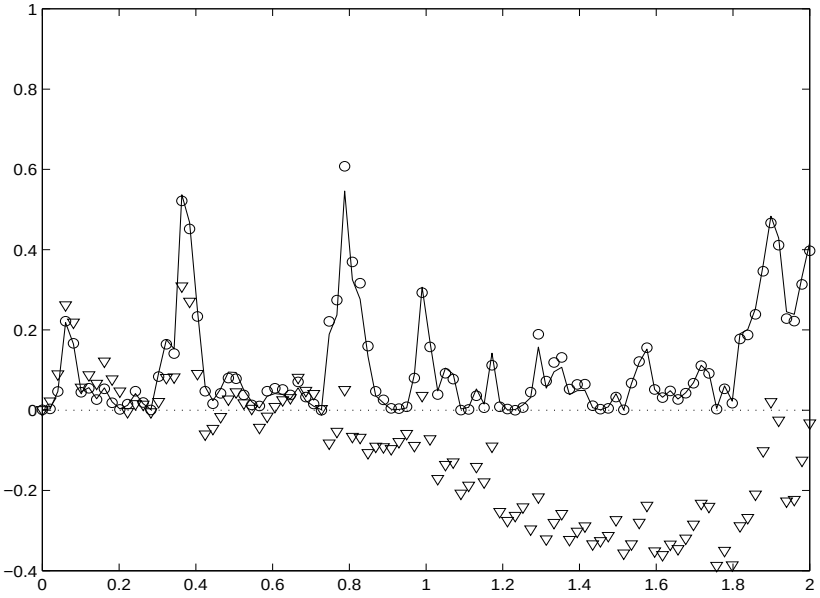



Figure 5.12: The dotted graph (o) depicts a simulated trajectory calculated by means of the scheme (5.40). The solid graph (-)corresponds to the Kloeden scheme (5.41) for the SDE (5.39) with s=0.01. Also included is a simulation from the Euler-Maruyama scheme ($\triangledown$).

# Chapter 6

# Weak solutions

## 6.1 Weak solution

The concept of a weak solution to a stochastic differential equation, is very different from that of a strong solution. Based on a *fixed* probability space, and a *fixed* Wiener process on that space, a strong solution may be found under suitable circumstances, using that particular Wiener process, and that particular probability space.

For a weak solution however, the probability space, the Wiener process, and the solution process, are themselves part of the solution of the stochastic differential equation.

**Definition 6.1.1** *The triple* $\big((X,W),(\Omega,\mathcal{F},\mathbb{P}),\{\mathcal{F}_t\}_{t\geq 0}\big)$ *is a* weak solution *to the one dimensional stochastic differential equation*

$$dX_t = a(t,X_t)\,dt + b(t,X_t)\,dW_t, \tag{6.1}$$

*if the following conditions hold:*

1. $(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\geq 0},\mathbb{P})$ *is a filtered, right-continuous and augmented probability space;*

2. $W$ *is a one dimensional Wiener process;*

3. $X$ *is a continuous and adapted stochastic process;*

4. $\forall t\in[0,\infty):\quad \mathbb{P}\Big(\int_0^t |a(s,X_s)|\,ds + b^2(s,X_s)\,ds < \infty\Big) = 1;$

5. $\mathbb{P}\Big(\forall t\in[0,\infty):\ X_t = X_0 + \int_0^t a(s,X_s)\,ds + \int_0^t b(s,X_s)\,dW_s\Big) = 1.$

## 6.2 The Feynman-Kac formula

Consider the SDE

$$\begin{aligned} dX(s) &= a(s, X(s))dt + b(s, X(s))\, dW_s \\ X(t) &= x \\ X(s), W_s &\in \mathbb{R}^n \\ s &\in [t, T] \end{aligned} \tag{6.2}$$

with coefficients that satisfy a Lipschitz condition

$$\forall s \in [t, t+T],\ \forall x, y \in \mathbb{R}: \quad |a(s,x) - a(s,y)| + |b(s,x) - b(s,y)| \leq K|x-y|.$$

Associated with the SDE (6.2) is the Kolmogorov backward partial differential operator, $\mathcal{A}$, called the *generator* of the SDE. It is given by

$$\mathcal{A} = \sum_{i=1}^{n} a_i(t,x)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(t,x)\frac{\partial^2}{\partial x_i x_j}, \tag{6.3}$$

where

$$c(s,x) = b(s,x)b(s,x)'.$$

Here $b(s,x)'$ is the transpose of the vector $b(s,x) = \big(b_1(s,x), \ldots, b_n(s,x)\big)$.

Using the generator $\mathcal{A}$, the Itô formula may be expressed as

$$df\big(t, X(t)\big) = \Big(\frac{\partial f}{\partial t} + \mathcal{A}f\Big)\big(t, X(t)\big)\, dt + (\nabla_x f \cdot b)\big(t, X(t)\big)\, dW_t,$$

where

$$\nabla_x f = \Big(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\Big)$$

and

$$\nabla_x f \cdot b = \frac{\partial f}{\partial x_1} b_1 + \ldots + \frac{\partial f}{\partial x_n} b_n.$$

Consider a solution $f(t,x)$ to the partial differential equation (PDE)

$$\begin{aligned} \frac{\partial f}{\partial t} + \mathcal{A}f &= rf, \quad \text{where } r \text{ is a constant} \\ f(T,x) &= \phi(x) \end{aligned} \tag{6.4}$$

which satisfies the following additional assumption

$$\int_t^T \mathbb{E}\big[(\nabla_x f \cdot b)^2(s, X(s))\big]\, ds < \infty.$$

With this framework, the solution to the PDE (6.4) can be represented, according to the Feynman-Kac formula, as

$$f(t,x) = \mathrm{e}^{-r(T-t)}\,\mathbb{E}[\phi(X(T))|X(t)=x]. \tag{6.5}$$

**Example 6.2.1** The task is to solve the PDE

$$\begin{aligned} \frac{\partial f}{\partial t} + \frac{1}{2}b^2\frac{\partial^2 f}{\partial x^2} &= 0 \\ f(T,x) &= I_{(-\infty,0]}(x) \end{aligned} \tag{6.6}$$

where $b$ is a constant. The corresponding SDE is

$$\begin{aligned} dX(s) &= b\,dW_s,\ s\in[t,T] \\ X(t) &= x \end{aligned}$$

with Itô-solution $X(T) = x + b(W_T - W_t)$. According to the Feynman-Kac formula (6.5), the solution to the PDE (6.6) is

$$\begin{aligned} f(t,x) &= \mathbb{E}\big[I_{(-\infty,0]}(X(T))\big|X(t)=x\big] \\ &= \mathbb{P}\big(X(T)\le 0\big|X(t)=x\big) = 1-\Phi\left(\frac{x}{\sqrt{\sigma^2(T-t)}}\right), \end{aligned}$$

where $\Phi$ is the standard Gaussian probability distribution function.

**Remark 6.2.1** The transition probability distribution function

$$\mathbb{P}\big(X(T)\le y\big|X(t)=x\big)$$

is obtained as the solution to the same PDE (6.6), with final value

$$f(T,x) = I_{(-\infty,y]}(x).$$

The above example shows not only how a solution to a PDE may be computed via an SDE, but also that the transition probability for the solution of the SDE may be computed as the solution of the PDE.

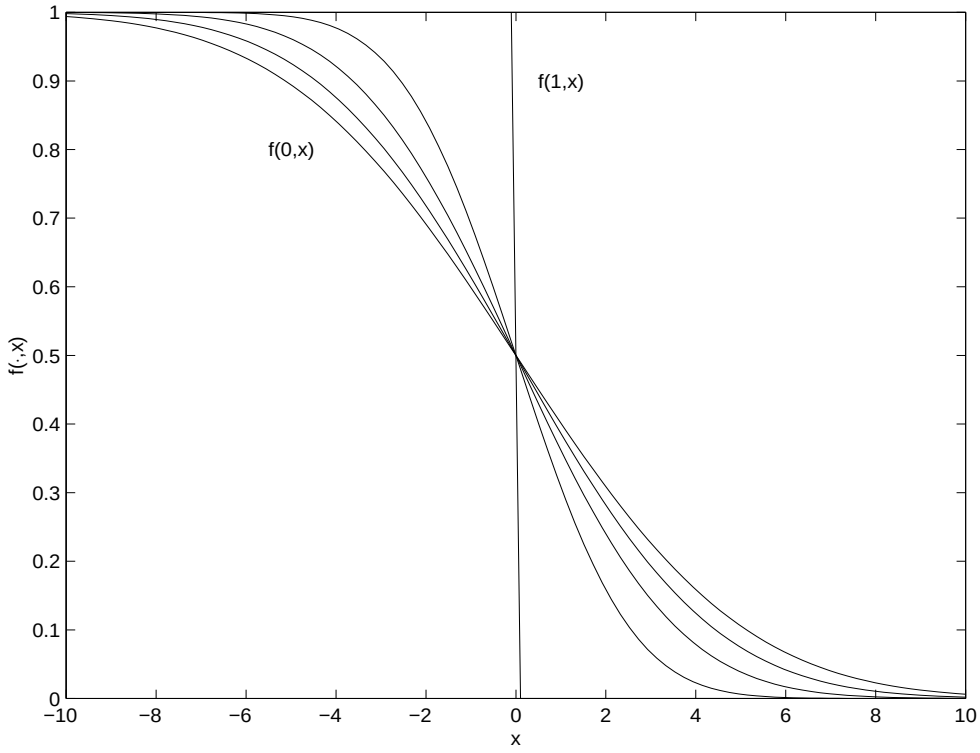



Figure 6.1: The solution $f(t,x)$ for $t \in \{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$, $x \in [-10, 10]$ and $b = 4$.

# Chapter 7

# Weak approximations

The expected value $\mathbb{E}[\phi\big(X(T)\big)|X(t)=x]$ in the Feynman-Kac formula (6.5) can be approximated by

$$\mathbb{E}[\phi\big(X(T)\big)|X(t)=x] \approx \frac{1}{M}\sum_{i=1}^{M}\phi\big(X_{i,(t,x)}(T)\big).$$

Here $\{X_{i,(t,x)}(T)\}_{i=1}^{M}$ are $M$ independent trajectories of the *exact* solution, evaluated at the final time $T$, given that the value at time $t<T$ is $x$.

The exact solution is seldom available, and therefore has to be approximated. Let $\overline{X}$ denote this approximation. In order to compute the solution of the PDE (6.4), two approximations have to be made:

$$\mathbb{E}[\phi\big(X(T)\big)|X(t)=x] \approx \mathbb{E}[\phi\big(\overline{X}(T)\big)|\overline{X}(t)=x] \approx \frac{1}{M}\sum_{i=1}^{M}\phi\big(\overline{X}_{i,(t,x)}(T)\big).$$

Thus $X_i(T)$ is approximated by a random variable $\overline{X}_i(T)$. In order for the approximation to be good, it is required that $X_i(T)$ and $\overline{X}_i(T)$ are close in the weak sense.

Let the set $\mathcal{H}$ contains "sufficiently many functions", as specified by Milstein [41]. Let $t=t_0<t_1<\ldots<t_N=T, t_{i+1}-t_i=h$ be a uniform partition of the interval $[t,T]$.

The approximation $\overline{X}$ is said to be of *weak order* $p$ if, for all $f\in\mathcal{H}$ and for all $k\in\{0,\ldots,N\}$,

$$\Big|\mathbb{E}\big[f(X(t_k))\big]-\mathbb{E}\big[f(\overline{X}(t_k))\big]\Big| \le Ch^p.$$

Here $C=C(a,b,x)$ is a constant that may depend only on the functions $a$ and $b$ and the initial value $x$ for the PDE (6.4).

According to Milstein, the first time a weak approximation of a stochastic differential equation was used, was in his own article [40] from 1978.

The main idea with weak approximations, is to construct methods that do not require the complicated modelling of complicated random variables, like multiple Itô integrals or Stratonovich integrals. Instead, these are replaced with simpler objects, that share some of their properties, like their first and second moments. For example, instead of using a Wiener process increment, i.e. a $N(0,h)$ distributed random variable, one may use a discrete random variable $V$ with the two-point distribution

$$\mathbb{P}(V=\sqrt{h})=\frac{1}{2} \quad \text{and} \quad \mathbb{P}(V=-\sqrt{h})=\frac{1}{2}.$$

The first two moments of this random variable

$$\mathbb{E}[V]=0 \quad \text{and} \quad \mathbb{E}[V^2]=h$$

coincide with those of the $N(0,h)$ distribution.

The following theorem, that can be found e.g. in Milstein's book [41], gives the fundamental connection between a one-step approximation and an approximation on a finite time interval. The theorem makes precise the following rule of thumb: "If the mean-square error in one time step, from $t_k$ to $t_{k+1}$, is of order $h^{p_2}$, then the mean-square error over the whole interval $[t_k, t_{k+1}]$, is of order $h^{p_2-1/2}$, provided that $p_2+\frac{1}{2}\leq p_1$", with $p_1$ given in the theorem:

**Theorem 7.0.1** *Consider the SDE* (6.2). *Let* $X_{t_0,x}(t)$ *be the exact Itô solution of* (6.2) *with* $X(t_0)=x$. *Consider the one-step approximation* $Y_{t_0,x}(t_0+h)$ *of* $X_{t_0,x}(t_0+h)$ *defined by*

$$Y_{t_0,x}(t_0+h)=x+F(t_0,x,h,\{W_i(\theta)-W_i(t_0)\}_{i=1}^n),$$

*for* $\theta\in[t_0,t_0+h]$, *and for some map* $F$ *specific for the approximation method used. Assume that, for every* $t\in[t_0,T-h]$ *and every* $x\in\mathbb{R}^n$,

$$\begin{aligned}
\|X_{t,x}(t+h)-Y_{t,x}(t+h)\|_1 &\leq K\sqrt{(1+|x|^2)}h^{p_1}\\
\|X_{t,x}(t+h)-Y_{t,x}(t+h)\|_2 &\leq K\sqrt{(1+|x|^2)}h^{p_2}\\
\|X_{t,x}(t+h)-Y_{t,x}(t+h)\|_4 &\leq K\sqrt[4]{(1+|x|^4)}h^{p_2-1/4}
\end{aligned}$$

*where*

$$\frac{5}{4}\leq p_2+\frac{1}{2}\leq p_1$$

*and*

$$\|Z\|_p = (\mathbb{E}|Z^p|)^{\frac{1}{p}}.$$

*Then it holds that*

$$\sqrt{\mathbb{E}\Big[\max_{0\le k\le N}\left|X_{t_0,X_0}(t_k) - \overline{X}_{t_0,X_0}(t_k)\right|^2\Big]} \le K\sqrt[4]{(1+\mathbb{E}\big[\,|X_0|^4\,\big]}\,h^{p_2-1/2},$$

*where* $t_{k+1} - t_k = h$, $t_N = T$ *and* $X_0 = X(t_0)$ *is a random initial value such that* $\mathbb{E}\big[\,|X_0|^4\,\big] < \infty$.

## 7.1 The probabilistic method and deterministic methods

A justified question to ask at this point, is whether the discussed probabilistic method based on the Feynman-Kac representation (6.5) is worthwhile, given the fact that there are many well developed deterministic methods for solution of PDE:s.

The probabilistic representation along with the Monte Carlo method might be worthwhile if the dimension of the PDE is so large, greater than four for the finite element method, that the deterministic algorithms become too computer intensive. The suitability of the representation also depends on the coefficients $a(s,x)$ and $b(s,x)$ in the generator (6.3).

Concerning the *computational cost*, it is to be noted that the deterministic algorithms have a cost that grows *exponentially* with the dimension, whereas the Monte Carlo method has a cost that grows only *linearly* with the dimension of the PDE.

The Monte Carlo method is an attractive alternative, if the solution of the PDE is sought at a few points only, and not over the entire range.

There is also a possibility of combining the Monte Carlo method with, for example, the finite element method using *domain decomposition techniques*. To present the idea, consider the solution of the PDE

$$\begin{aligned}\frac{\partial f}{\partial t} + \mathcal{A}f &= 0\\ f(T,x) &= \phi(x)\\ (t,x) &\in [0,1]\times[0,1]\end{aligned}$$

The domain $[0,1]\times[0,1]$ is divided into, for example, four sub-domains. At the boundaries of these domains, the solution is approximated by the Monte Carlo method. Then the deterministic method is used on each domain to compute the solution to the PDE, with the Dirichlet boundary conditions

supplied by the Monte Carlo method.

For further information, consult Talay's article in the proceedings [57].

### 7.1.1 The waveform relaxation method

If the dimension of the problem is very large, say 1000 or more, then the waveform relaxation method for SDE might be considered, since this method is specifically designed to handle very large systems of SDE:s. It is however a requirement that the components of the solution vector are weakly interacting in order for the waveform relaxation method to work. There might also be some problems with convergence, but there have been designed methods that accelerate the convergence for specific cases.

## 7.2 Research on weak numerical methods

This chapter on weak numerical methods is considerably shorter than that on strong numerical methods. The major reason for this is that the material on weak numerical methods seems to be relatively sparse, compared to that on strong methods. This has to do with the fact that many strong methods are adaptations of methods from the well-researched area of numerical methods for ordinary differential equations.

# Chapter 8

# Probability densities

## 8.1 The Fokker-Planck-Kolmogorov equation

In great generality, the transition probability densities

$$p(x, y, s, t) = \frac{\partial}{\partial y}\mathbb{P}\big(X(t) \leq y \big| X(s) = x\big)$$

of a solution $X(t)$ to the stochastic differential equation (6.2) satisfies a partial differential equation. That PDE is called the Fokker-Planck-Kolmogorov (FPK) equation, and is given by

$$\frac{\partial p(x, y, s, t)}{\partial t} - \mathcal{A}p(x, y, s, t) = 0.$$

Here, $\mathcal{A}$ is the generator given by (6.3). To find the evolution of the transition probabilities thus amounts to solve a PDE.

The Fokker-Planck-Kolmogorov equation governs the probability distribution of the response of a dynamical system subject to random excitations.

The two main objectives when solving the FPK equation are the stationary solution and the non-stationary solution. See [56] on numerical methods.

The one-dimensional Fokker-Planck equation is a linear second order parabolic partial differential equation:

$$\frac{\partial p}{\partial t} = -D^{(1)}(x, t)\frac{\partial p}{\partial x} + D^{(2)}(x, t)\frac{\partial^2 p}{\partial x^2},$$

where $D^{(1)}$ and $D^{(2)}$ are respectively the drift coefficient and the diffusion coefficient. In particular, by integration, the FPK equation determines both stationary and transient moments of arbitrary orders.

Exact solutions to the FPK equation can be found under the following circumstances [49]:

The drift and diffusion coefficients are linear, in which case both the transient and stationary probability distributions are Gaussian.

The stationary solution to the one dimensional FPK equation can be solved by quadratures.

If separation of variables in multidimensional FPK is possible

In the case of a multi-dimensional FPK equation, as a rule, an exact solution is impossible to find. This makes numerical approximations necessary.

The solution of FPK equation can be approximated, for example, in the following ways:

By power series expansion.

By Edgeworth expansion.

**Power series expansion** Consider the SDE

$$dX_i(t) = a_i(X(t),t)\,dt + \sum_{j=1}^{n} b_{ij}(X(t),t)\,dW_j(t), \quad i = 1,\ldots,n. \tag{8.1}$$

If the diffusion coefficient $b(x,t)$ is constant, and the drift coefficient $a(x,t)$ can be represented as a power series, then the solution of the FPK equation can be expressed as

$$\log p(x_1,\ldots,x_n,t) =$$
$$A_0(t) + \sum_{i=1}^{n} A_i(t)x_i + \frac{1}{2}\sum_{i,j=1}^{n} A_{ij}(t)x_i x_j + \frac{1}{3}\sum_{i,j,k=1}^{n} A_{ijk}(t)x_i x_j x_k + \ldots\ .$$

Here the coefficients $A_{(\cdot)}(t)$ are obtained by numerically solving a system of ordinary differential equations.

**Edgeworth expansion** Consider the SDE (8.1). If the drift coefficient $a(x,t)$ is linear, or "weakly non-linear", see [49], then the solution to the FPK equation can be represented as a power series, involving the standard normal distribution. There is a drawback with this method, namely the possibility that the approximated probability distribution takes on negative values. See Sobczyk and Trębicki [55] for more information on the use of power series expansions, and Edgeworth expansions, to approximate solutions of SDE:s.

Regarding the Fokker-Planck equation for stochastic Hamiltonian systems, Soize gives a thorough discussion in his book [56].

## 8.2 The Maximum-entropy method

This method works for large systems of SDE:s. An approximate analytical solution is obtained. One of the drawbacks is the requirement to evaluate several multiple integrals. This problem may be overcome by the use of fast Fourier transforms, as described in [21].

### 8.2.1 Classical maximum entropy method

The method uses a finite set of moments $\{m_k\}_{k=1}^n$ to approximate an unknown underlying probability distribution, $f$. The aim of the approximation is to maximise the so called *entropy*, $H(f)$, given by

$$H(f) = -\int f(x) \log f(x) dx.$$

The problem at hand is thus to maximise a certain function, under some imposed conditions. To find the maximum, if it exists, one augments the entropy with so called *Lagrange multipliers* $\lambda_i$. The augmented entropy function is given by

$$H(f, \lambda_1, \ldots, \lambda_n) = H(f) + \sum_{i=1}^{n} \lambda_i (m_k - \int x^k f(x) dx).$$

The solution to the maximisation problem is of the form

$$f_M(x) = C \exp\left( -\sum_{i=1}^{n} \lambda_i x^i \right), \tag{8.2}$$

where the Lagrange multipliers are determined by the constraints

$$\begin{aligned} \int f_M(x) dx &= 1 \\ \int x f_M(x) dx &= m_1 \\ &\vdots \\ \int x^n f_M(x) dx &= m_n \end{aligned}$$

This method performs well (when compared with others), for SDE of dimension *higher* than three.

### 8.2.2 Systems of stochastic differential equations

The classical maximum entropy method is applied to stochastic differential equations by Trębicki and Sobczyk in [58]. They consider the following system of stochastic differential equations

$$\begin{aligned} d\mathbf{X}_t &= \mathbf{F}(\mathbf{X})_t\, dt + \sigma(\mathbf{X}_t)\, d\mathbf{W}_t \\ \mathbf{X}_{t_0} &= \mathbf{X}_0 \end{aligned} \tag{8.3}$$

where

$$\begin{aligned} \mathbf{X}_t &= (X_{1,t}, \ldots, X_{n,t})' \\ \sigma &\in \mathbb{R}_{n|m} \end{aligned} \tag{8.4}$$

The probability distributions, of the solution process $\mathbf{X}$ for (8.3) and (8.4), cannot be expressed explicitly, with the exception for a few particular cases. Thus, one has to resort to other means to study probabilities for $\mathbf{X}$. One way is to consider the moments of $\mathbf{X}$.

Suppose that, for the solution $\mathbf{X}$ to (8.3) and (8.4), the moments

$$\begin{aligned} m_k(t) &= \mathbb{E}[h_k(\mathbf{X})] \\ h_k(\mathbf{X}) &= X_1^{k_1}(t) \cdot \ldots \cdot X_n^{k_n}(t) \\ k &= k_1 + \ldots + k_n \end{aligned}$$

are sought. By application of the multidimensional Itô formula, the evolution of the moments are governed by the following equation:

$$\begin{aligned} \frac{dm_k}{dt} &= \mathbb{E}[G_k(\mathbf{X}, t)] \\ m_k(t_0) &= m_{k_0} \\ G_k(\mathbf{X}, t) &= \sum_i F_i(\mathbf{X}, t) \frac{\partial h_k(\mathbf{X})}{\partial x_i} + \frac{1}{2} \sum_l \sum_{i,j} \sigma_{il}(\mathbf{X}, t) \sigma_{jl}(\mathbf{X}, t) \frac{\partial^2 h_k(\mathbf{X})}{\partial x_i \partial x_j} \end{aligned} \tag{8.5}$$

Notice that (8.5) represents an infinite hierarchy of moments, since $k = 1, 2, \ldots$. Thus, the numerical solution of (8.5) requires the imposition of closure. With some specified finite $K$, the problem then becomes

$$\begin{aligned} \frac{dm_k}{dt} &= \mathbb{E}[G_k(\mathbf{X}, t)] \\ m_k(t_0) &= m_{k_0} \\ k &= 1, 2, \ldots, K \end{aligned} \tag{8.6}$$

Let $p(\mathbf{x}, t)$ be the approximate probability density of the solution $\mathbf{X}(t)$ of the stochastic differential equation (8.3). In analogy with (8.2), $p(\mathbf{x}, t)$ takes

the following form:

$$\begin{aligned} \log p(\mathbf{x},t) &= \log C(t) - \sum_{k=1}^{K} \lambda_k(t) G_k(\mathbf{x},t) \\ \log C(t) &= -\lambda_0(t) - 1 \end{aligned} \tag{8.7}$$

To solve (8.7), it remains to find the Lagrange multipliers $\{\lambda_i(t)\}_{i=0}^{K}$. These can be determined by evaluating the moments in (8.6), under the approximate probability density $p(\mathbf{x},t)$.

### 8.2.3 Numerical solution

Consider a discretization of the system (8.6), using an Euler approximation,

$$m_k(t_{i+1}) = m_k(t_i) + \Delta t \int_{\mathbb{R}^n} G_k(\mathbf{x},t_i) p(\mathbf{x},t_i)\, d\mathbf{x}.$$

Here the calculation of the $n$-dimensional integral $\int_{\mathbb{R}^n} G_k(\mathbf{x},t_i)p(\mathbf{x},t_i)d\mathbf{x}$ is very computer intensive, as is discussed in the article [55] by Sobczyk and Trębicki. In order to alleviate this problem, Hurtado and Barbat [21] have suggested reformulating the integral, in terms of multidimensional Fourier transforms.

# Chapter 9

# Moments

## 9.1 The cumulant-neglect-closure method

If the characteristic function of a distribution is known, then it is possible to determine the moments of all orders. Conversely, if the moments of all orders are known, with some additional technical constraints, then the characteristic function can be obtained. In practice however, all the moments are usually not accessible, but only a finite number of them. Then it becomes a question of using these finite set of moments to approximate the characteristic function.

### 9.1.1 Statistical cumulants

Consider the following system of stochastic differential equations:

$$\frac{\partial \mathbb{X}}{\partial t} = m(\mathbb{X}) + \mathbb{G}(\mathbb{X})\mathbb{W}(t), \tag{9.1}$$

where

$$\begin{aligned}
\mathbb{X} &= (X_1, X_2, \ldots, X_n) \in \mathbb{R}^n \\
\mathbb{W} &= (W_1, W_2, \ldots, W_m) \in \mathbb{R}^m \\
\mathbb{G} &= \begin{pmatrix} G_{11} & G_{12} & \ldots & G_{1m} \\ G_{21} & G_{22} & \ldots & G_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ G_{n1} & G_{n2} & \ldots & G_{nm} \end{pmatrix} : \mathbb{R}^n \to \mathbb{R}_{n|m} \\
m &= (m_1, m_2, \ldots, m_n) : \mathbb{R}^n \to \mathbb{R}^n
\end{aligned}$$

The vector Wiener process $\mathbb{W}$ has the following properties:

$$\begin{aligned}
\mathbb{E}[\mathbb{W}(t)] &= 0 \\
\mathbb{E}[\mathbb{W}(t_1)\mathbb{W}(t_2)'] &= 2\mathbb{D}\delta(t_1 - t_2) \\
\mathbb{D} &= \begin{pmatrix} D_{11} & 0 & \dots & 0 \\ 0 & D_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & D_{nn} \end{pmatrix} \in \mathbb{R}_{n|n} \\
\mathbb{W}(t)' &= \begin{pmatrix} W_1(t) \\ \vdots \\ W_m(t) \end{pmatrix}
\end{aligned}$$

The aim is to obtain the moments of the solution process $\mathbb{X}$ to (9.1). To that end, the function $g(x) = \prod_{i=1}^{n} x_i^{s_i}$ is formed, where $s_i \in \mathbb{N}$.

The Itô formula is applied to $g(\mathbb{X})$ in order to obtain an equation governing the evolution of the moments, $\alpha(s_1, \dots, s_n) = \mathbb{E}[g(\mathbb{X}(t))]$:

$$\begin{aligned}
&\frac{d\alpha}{dt}(s_1, \dots, s_n) \\
&= \sum_{i=1}^{n} \mathbb{E}\Big[ m_i(\mathbb{X}(t)) \frac{\partial g}{\partial x_i}(\mathbb{X}(t)) \Big] + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\Big[ H_{ij}(\mathbb{X}(t)) \frac{\partial^2 g}{\partial x_i \partial x_j}(\mathbb{X}(t)) \Big].
\end{aligned}$$

Here $H_{ij} = (2\mathbb{G}\mathbb{D}\mathbb{G}')_{ij}$.

For example, if the first moment of component $X_1$ is sought, then we take $g(x) = x_1$, and the equation becomes:

$$\frac{d\alpha}{dt}(s_1, 0, \dots, 0) = \frac{d\mathbb{E}[X_1(t)]}{dt} = \mathbb{E}\big[ m(\mathbb{X}(t)) \big].$$

Notice that already this first order moment may depend on the solution $\mathbb{X}$ to (9.1) in a complicated manner, and in particular, on moments of higher order, depending on the form of the function $m_1$.

From the above example, it seems that the problem of determining the moments is a highly complex one. Especially, this is so for large dimensions $n$, as is usually the case for dynamical systems occurring in engineering applications. It is also possible that financial applications may involve several interacting processes $X$. For example, the stock prices in telecommunications may be related to each other, thus forming a complicated stochastic dynamical system. The point that is being made here is that high-dimensional processes are important to describe real world phenomena, thus meriting a study.

With this we come to the subject of this section, namely the so called *cumulant neglect closure method*, which is discussed by Bergman, Wojtkiewicz and Johnson [2] and Wojtkiewicz, Spencer and Bergman [61]. This method has been devised to reduce the complexity of the problem to determine the moments of a stochastic dynamical system. Basically, it is a truncation of the logarithm of the characteristic function $\phi_{\mathbb{X}}$, of the vector process $\mathbb{X}$. The truncation is made assuming that the so called *response cumulants* above a certain threshold can be neglected, in comparison with response cumulants below that threshold.

The characteristic function $\phi_X(\theta)$ for a one-dimensional random variable can be represented by the following exponential map

$$\phi_X(\theta) = \exp\left(\sum_{n=0}^{\infty} \kappa_n(X)\frac{(i\theta)^n}{n!}\right).$$

Here $\kappa_n(X)$ are the so called *cumulants of order* $n$ for the random variable $X$, and are given by

$$i^n\kappa_n(X) = \frac{d^n \log \phi_X(\theta)}{d\theta^n}\Big|_{\theta=0}.$$

Cumulants become more useful in the context of vector-valued random variables. Naturally, the corresponding formulae then become more involved.

### 9.1.2 Taylor cumulants

Expand the characteristic function $\phi_X(\theta)$ of a random variable $X$ in a Taylor series

$$\phi_X(\theta) = \sum_{k=0}^{\infty} c_k\theta^k, \tag{9.2}$$

where

$$c_k = \frac{\mathbb{E}[X^k]}{k!}. \tag{9.3}$$

If the random variable $X$ that is under consideration in (9.2) is the solution $X_t$ to the SDE

$$dX_t = a(X_t, t)\,dt + b(t)\,dW_t, \tag{9.4}$$

then the moments $m_k(t) = \mathbb{E}[X_t^k]$ of $X = X_t$ that feature in (9.3), can be obtained from the following system of differential equations:

$$\begin{aligned}
\frac{dm_1(t)}{dt} &= \mathbb{E}[a(X_t,t)] \\
\frac{dm_2(t)}{dt} &= \mathbb{E}[a(X_t,t)X_t] + b^2(t) \\
&\vdots \\
\frac{dm_k(t)}{dt} &= \mathbb{E}[a(X_t,t)X_t^{k-1}] + \frac{k(k-1)}{2}b^2(t)m_{k-2}(t) \\
&\vdots
\end{aligned} \tag{9.5}$$

Note that if the drift-coefficient, $a(X_t,t)$ in the equation (9.4) is a polynomial of degree 2 or more, then the system (9.5) cannot be expected to be solved explicitly, since the moment of order $k$ depends on moments of order higher than $k$.

An approximation, $\psi(\theta,t)$, to the characteristic function $\phi(\theta,t)$ in (9.2), when the random variable $X$ is the solution $X = X_t$ to the SDE (9.4), can be obtained by setting the moments of order greater than some fixed number $k_0$ equal to zero. This will reduce (9.5) to a finite system of equations, that can be solved. However, a possible severe shortcoming of this procedure, is that $\psi(\theta,t)$ need not be a characteristic function of a probability distribution. In order to overcome this problem, the corresponding system of cumulants, $\kappa_r(t)$, given by

$$\begin{aligned}
&\vdots \\
\frac{d\kappa_r(t)}{dt} &= i^{-1-3r}\frac{\partial^r}{\partial\theta^r}\left\{\frac{\theta}{\phi(\theta,t)}\,\mathbb{E}\big[\exp(-i\theta X_t)a(X_t,t)\big] + \frac{\theta^2}{2}b^2(t)\right\}\Bigg|_{\theta=0} \\
&\vdots
\end{aligned} \tag{9.6}$$

see [36], is studied instead.

Note that the cumulants are not given explicitly by the system of equations (9.6), and that this system is an infinite hierarchy just as is the system of moments (9.5). To obtain an approximation to $\phi(\theta,t)$, the cumulants of order larger than some fixed order $k_0$ are set equal to zero. However, due to the implicit nature of the resulting equations, the solution is still not easy to calculate. In addition, there is the problem that the equations are non-linear when the drift coefficient is non-linear.

The above outlined method will work for SDE with linear or linearly approximated drift. If the system (9.6) can be solved, the task of computing the probability density, $p(x)$, of the solution $X_t$, by means of inverse Fourier transform, still remains.

The indicated problems suggests that an alternative approach to compute the probability density $p(x,t)$ for the solution $X_t$ to the SDE (9.4) is needed. Such an approach has been suggested by DiPaola, Ricciardi and Vasta [36]. They start by Taylor expanding the *probability density* $p(x,t)$

$$p(x,t) = \sum_{k=0}^{\infty} \frac{\mu_k(t)}{k!} x^k \quad \text{where} \quad \mu_k(t) = \left. \frac{\partial^k p(x,t)}{\partial x^k} \right|_{x=0}.$$

The functions $\mu_k$ are related to the characteristic function $\phi(\theta,t)$ by

$$\mu_k(t) = \frac{1}{2\pi} \int_{\mathbb{R}} (i\theta)^k \phi(\theta,t)\, d\theta.$$

Due to their resemblance to moments, these functions are called *Taylor moments*.

An alternative representation of $p(x,t)$ is

$$\log p(x,t) = \sum_{k=0}^{\infty} \lambda_k(t) x^k.$$

This representation assures the positivity of any truncation

$$p(x,t) \approx \tilde{p}(x,t) = \exp\left( \sum_{k=0}^{n} \lambda_k(t) x^k \right).$$

Here

$$\lambda_k(t) = \left. \frac{\partial^k \log p(x,t)}{\partial x^k} \right|_{x=0}$$

are called the *Taylor cumulants*.

The SDE (9.4) generates the following system of differential equations for the Taylor cumulants:

$$\begin{aligned} &\vdots \\ \frac{d\lambda_k(t)}{dt} &= \eta_k(t) + \beta(t)\lambda_{k+2}(t) + \sum_{i=0}^{k} \gamma_{ki}(t)\lambda_{k+1}(t) + \beta_{ki}(t)\lambda_{k+1}(t)\lambda_{k-i+1}(t) \\ &\vdots \end{aligned} \tag{9.7}$$

where

$$\eta_k(t) = -\frac{\partial^{k+1} a(x,t)}{\partial x^{k+1}}\bigg|_{x=0}$$
$$\beta(t) = \frac{b^2(t)}{2}$$
$$\beta_{ki}(t) = \binom{k}{i}\beta(t)$$
$$\gamma_{ki}(t) = -\binom{k}{i}\frac{\partial^{k-i} a(x,t)}{\partial x^{k-i}}\bigg|_{x=0}$$

Again an infinite hierarchy of equations feature, so that a closure scheme has to be used. As before, Taylor cumulants of order greater than some fixed $k_0$ are set to zero, and the resulting finite system of equations is solved.

The following things are worth noting for the above described method:

Taylor cumulants appear explicitly in the equations (9.7) instead of implicitly as the cumulants in equations (9.6).

The positivity of the approximate probability distribution is assured.

Taylor cumulants translate directly to the approximate probability distribution. (No inverse Fourier transformation is required.)

The method using Taylor cumulants works also for a non-linear drift $a(x,t)$.

The method described also works well for the multidimensional SDE

$$\begin{cases} dX_i &= f_i(X_1,\ldots,X_n,t)\,dt + \sum_j G_{ij}(t)\,dW_j(t) \\ X_i(0) &= X_{0i} \end{cases}$$

The infinite hierarchy of differential equations becomes very complex in this situation, and is therefore not presented here. However, the basic ideas, from the one dimensional setting, remain the same.

In their article [35], Lutes and Papadimitriou present an alternative way of formulating the system of differential equations, governing the evolution of the cumulants of a stochastic dynamical system. Their proposed method applies to both linear and non-linear systems.

# Bibliography


[1] B G Arous. Flots et series de Taylor stochastiques. *Probab. Theory related Fields*, 81:29–77, 1989.

[2] L A Bergman, S F Wojtkiewicz, and E A Johnson. Some reflections on the efficacy of moment closure methods. In *Second international conference on Computational Stochastic Mechanics*, June 1994.

[3] S. N. Bernstein. Principes de la theorie des equations differentielles stochastiques. *Trudy Fiz.-Mat. Inst. Akad. Nauk*, 5:95–124, 1934.

[4] T Björk. *Arbitrage theory in continuous time.* Oxford University Press, 1998.

[5] K Burrage and P M Burrage. High strong order methods for non-commutative stochastic ordinary differential equation systems and the Magnus formula. *Physica D*, 133:34–48, 1999.

[6] J C Butcher. *The numerical analysis of ordinary differential equations.* Wiley interscience, 1987.

[7] P J Channell and C Scovel. Symplectic integration of Hamiltonian systems. *Nonlinearity*, 3:231–259, 1990.

[8] P.E. Crouch and R. Grossman. Numerical integration of ordinary differential equations on manifolds. *J. Nonlinear Sci.*, 3(1):1–33, 1993.

[9] L Debnath and P Mikusinski. *Introduction to Hilbert spaces with applications.* Academic Press, 1998.

[10] K J Falconer. *Geometry of Fractal Sets.* Cambridge University Press, 1985.

[11] J G Gaines and T J Lyons. Variable step size control in the numerical solution of stochastic differential equations. *SIAM Journal of Applied Mathematics*, 57(5):1455–1484, October 1997.

[12] T Gard. *Introduction to stochastic differential equations.* Marcel-Dekker, New York, 1988.

[13] H. Geman and M. Yor. Bessel processes, Asian options and perpetuities. *Math. Finance*, 3:349–375, 1993.

[14] E. Hairer. Backward error analysis of numerical integrators and symplectic methods. In *Proceedings of the SCADE'93 conference, Auckland, New Zealand*, volume 1, pages 107–132, January 1993.

[15] E. Hairer and C. Lubich. The life-span of backward error analysis for numerical integrator. *Numer.Math.*, 76:441–462, 1997.

[16] F. Harary. *Graph Theory.* Addison-Wesley, Reading MA, 1969.

[17] F. Hausdorff. Die symbolische exponentialformel in der gruppentheorie. *Berichte der Sächsischen Akademie der Wissenschaften (Math. Phys. Klasse)*, 58:19–48, 1906.

[18] S Helgason. *Differential geometry, Lie groups and Symmetric spaces.* Academic Press, 1978.

[19] C.C Heyde. *Quasi-likelihood and its application: A general approach to optimal parameter estimation.* Springer, 1997.

[20] N Hofmann, T Müller-Grönbach, and K Ritter. The optimal discretization of stochastic differential equations. *Journal of Complexity*, 17:117–153, March 2001.

[21] J E Hurtado and A H Barbat. Fourier-based maximum entropy method in stochastic dynamics. *Structural Safety*, 20:221–235, 1998.

[22] N Ikeda and S Watanabe. *Stochastic differential equations and diffusion proceses.* North Holland, Amsterdam, 1981.

[23] A Iserles et al. Lie-group methods. In *Acta Numerica*, volume 9, pages 215–365. Cambridge University Press, 2000.

[24] A. and Iserles. On the solution of linear differential equations in Lie groups. *Phil. Trans. Royal Society A*, 357:983–1020, 1999.

[25] B. Jacob. *Linear Algebra.* W.H. Freeman and company, 1990.

[26] J. Janssen and S. Vandewalle. Convolution-based chebyshev acceleration of waveform relaxation methods. Technical report, Katholieke Universiteit Leuven, dept. Computerwetenschappen, 2001.

[27] R Kannan and C K Krueger. *Advanced analysis on the real line.* Springer-Verlag, 1996.

[28] I Karatzas and S E Shreve. *Brownian motion and Stochastic calculus.* Number 113 in Graduate texts in mathematics. Springer, 1991.

[29] P E Kloeden and E Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics.* Springer-Verlag, 1992.

[30] H J Kushner. *Probability methods for approximations in stochastic control and for elliptic equations.* Academic Press, 1977.

[31] B. Leimkuhler. Timestep acceleration of waveform relaxation. *SIAM Journal on Numerical Analysis*, 35(1):31–50, 1998.

[32] E Lelarasmee. *The Waveform relaxation method for the time domain analysis of large scale non-linear systems.* PhD thesis, University of California, Berkeley, 1982.

[33] P Lévy. Processus Stochastiques et Mouvement Brownien. *Monographies des Probabilités*, 1948.

[34] A Logg, C Johnson, and K Eriksson. Explicit time-stepping for stiff ODE:s. Preprint, Chalmers Finite Element Center, July 2002.

[35] L D Lutes and C Papadimitriou. Direct derivation of response moment and cumulant equations for non-linear stochastic problems. *International journal of Non-linear mechanics*, 35:817–835, 2000.

[36] Di Paola M, Ricciardi G, and Vasta M. A method for the probabilistic analysis of nonlinear systems. *Probabilistic Engineering Mechanics*, 10:1–10, 1995.

[37] G Maruyama. Continuous Markov processes and stochastic equations. *Rend. Circolo Math. Palermo*, 4:48–90, 1955.

[38] P Mattila. *Geometry of sets and measures in Euclidean spaces : Fractals and rectifiability.* Cambridge University Press, 1995.

[39] S.D. Miller and A.G. Miller. Multidimensional toroidal compactification with random gaussian modulus: Derivation and potential applications of a stochastic general relativity. *General Relativity and Gravitation*, 33(6):937–964, 2001.

[40] G A Milstein. A method of second order of accuracy for integrating stochastic differential equations. *Teor. Veroyatn. i Prilozhen*, 23(2):414–419, 1978.

[41] G A Milstein. *Numerical Integration of Stochastic Differential Equations*. Cambridge University Press, 1990.

[42] G N Milstein, Yu M Repin, and M V Tretyakov. Symplectic integration of Hamiltonian systems with additive noise. *SIAM Journal of Numerical Analysis*, 39(6):2066–2088, 2002.

[43] G N Milstein and Tretyakov M V. Numerical methods for Langevin type equations based on symplectic integrators. Technical report, Weierstrass-Institut für Angewandte Analysis und Stochastik, 2000.

[44] T Misawa. Lie alebraic approach to numerical integration of stochastic differential equations. *SIAM J. Sci. Comput.*, 23(3):866–890, 2001.

[45] P.C. Moan. Efficient approximation of sturm-liouville problems using lie-group methods. Technical Report NA11, DAMTP, University of Cambridge, 1998.

[46] A Neumaier. Molecular modeling of proteins and methematical prdiction os protein structure. *SIAM Review*, 39(3):407–460, 1997.

[47] A.L. Onishchik and E.B. Vinberg. *Lie groups and Algebraic groups*. Springer-Verlag, 1990.

[48] E. Platen and W. Wagner. Approximaton of ito integral equations. *Preprint ZIMM, Akad. der Wiss. der DDR, Berlin*, 1978.

[49] H Risken. *The Fokker-Planck equation: Methods of solution and applications*. Springer-Verlag, 1989.

[50] T Ryden and M Wiktorsson. On the simulation of iterated Ito integrals. *Stochastic Processes and their Applications*, 91:151–168, 2001.

[51] J M Sanz-Serna and M P Calvo. *Numerical Hamiltonian problems*. Chapman-Hall, 1994.

[52] K R Schneider and H Schurz. Waveform relaxation methods for stochastic differential equations, December, 5th 2000. Preprint at Weierstrass-Institut fur Angewandte Analysis und Stochastik.

[53] L.F. Shampine. *Numerical solution of ordinary differential equations.* Chapman and Hall, 1994.

[54] K Sobczyk. *Random Fatigue - From data to theory.* Academic Press, Boston, 1992.

[55] K Sobczyk and J Trębicki. Approximate probability distributions for stochastic systems: Maximum entropy method. *Computer methods in applied mechanics and engineering*, 168:91–111, 1999.

[56] C Soize. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions*, volume 17 of *Series on Advances in Mathematics for Applied sciences.* World Scientific, 1994.

[57] D. Talay and L. Tubaro, editors. *Probabilistic Models for Nonlinear Partial Differential Equations*, Lecture Notes in Mathematics. Springer, 1995.

[58] J Trębicki and K Sobczyk. Maximum entropy principle and non-stationary distributions of stochastic systems. *Probabilistic Engineering Mechanics*, 11:169–178, 1996.

[59] S Vanderwalle. *Parallel Multigrid Waveform Relaxation for Parabolic Problems.* Teubner-Skripten zur Numerik. B. G. Teubner, Stuttgart, 1993.

[60] V S Varadarjan. *Lie groups, Lie Algebras and their applications*, volume 102 of *Graduate texts in mathematics.* Springer-Verlag, 1984.

[61] S F Wojtkiewicz, Jr B F Spencer, and L A Bergman. On the Cumulant-Neglect Closure method in Stochastic Dynamics. *Int. J. Non-linear Mechanics*, 31:657–684, 1996.

[62] E Wong and M Zakai. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Statist.*, 36:1560–1564, 1965.

[63] E Wong and M Zakai. On the relation between ordinary and stochastic differential equations. *Int. J. Engng. Sci.*, 3:213–229, 1965.

[64] T Yamada and S Watanabe. On the uniqeness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11:155–167, 1971.