# Contents

1 Introduction and preliminaries 3  
1.1 Population models .................................................. 3  
1.2 Pricing of stock options ........................................... 4  
1.3 Stochastic optimal control ....................................... 5  
1.4 Probability and measure theory .................................. 5  
1.5 The Lebesgue integral ............................................. 8  
1.6 Expectation .......................................................... 10  
1.7 Conditional expectation .......................................... 12  
1.8 Convergence of random sequences .............................. 13  

2 Stochastic processes 13  
2.1 Markov and diffusion processes ................................. 15  
2.2 Brownian motion .................................................... 16  

3 Stochastic integrals 18  
3.1 Background ........................................................ 18  
3.2 The Itô integral ..................................................... 19  
3.3 Martingales .......................................................... 26  
3.4 Itô’s formula ......................................................... 28  
3.5 Vector valued stochastic differentials .......................... 32  
3.6 The Stratonovich integral ....................................... 33  

4 Stochastic differential equations 34  
4.1 Analytical solutions ............................................... 34  
4.2 Existence and uniqueness of solutions .......................... 38  
4.3 Strong and weak solutions ...................................... 45  

5 Connecting stochastic and partial differential equations 46  
5.1 Itô diffusions ......................................................... 46  
5.2 The Kolmogorov equations and Feynman-Kac’s formula .... 51  
5.3 The Black-Scholes equation and application to finance .... 55  

6 Stochastic optimal control ........................................ 57
7 An introduction to numerical approximation
  7.1 Euler approximation .............................................. 63
  7.2 Strong and weak approximation .................................. 64
  7.3 Statistical error and time discretization error ................. 65
  7.4 Consistency and convergence .................................... 68
  7.5 Strong order of convergence for the Euler approximation ........ 69
  7.6 Weak order of convergence for the Euler approximation ........ 72
  7.7 Stochastic numerical stability ................................... 74

8 Itô-Taylor expansion ................................................. 76

9 Strong approximation .................................................. 81
  9.1 Milstein approximation ........................................... 81
  9.2 Higher-order strong Itô-Taylor approximation .................... 81
  9.3 Explicit strong approximation ................................... 84
  9.4 Implicit strong approximation ................................... 85

10 Weak approximation .................................................. 85
  10.1 Higher-order weak Itô-Taylor approximation ..................... 86
  10.2 Explicit weak approximation .................................... 87
  10.3 Implicit weak approximation and predictor-corrector methods .... 88
  10.4 Extrapolation methods ........................................... 89

11 Adaptive weak approximation ....................................... 90
  11.1 Adaptive approximation with deterministic time steps ........... 91
  11.2 Adaptive approximation with stochastic time steps ............. 94

12 Variance reduction ................................................... 96
  12.1 Girsanov transformation .......................................... 96
  12.2 Variance reduced estimator ....................................... 99
1 Introduction and preliminaries

The aim of this course is to give a basic understanding of the theory of stochastic differential equations (SDE). The theory of stochastic differential equations is fairly new – in fact the first rigorous theory was published in 1951 – and the theoretical machinery required in order to define SDEs is quite heavy. As a result, we must begin the course by defining a large numbers of concepts, most of which are unfamiliar to readers that are not acquainted with axiomatic probability theory and the Lebesgue integral. When this theoretical machinery has been set up, the stochastic calculus should be rather straightforward. Most SDEs cannot be solved explicitly and, hence, a large part of the course will be devoted to finding reliable and efficient numerical methods for stochastic differential equations.

As a motivation, we shall first, in Sections 1.1-1.3, consider a couple of situations that can be modelled with stochastic differential equations.

1.1 Population models

A simple model for the size $N$ of a population at time $t$ is the model of exponential growth

$$\frac{dN}{dt} = rN, \quad (1.1)$$

where the growth parameter $r$ is often considered to be a positive constant. To model a population undergoing seasonal variations we can, however, let $r$ be non-constant with varying sign. One obvious drawback of this model is that, for strictly positive $r$, it predicts unlimited growth of the population. To avoid this unphysical behaviour, it is standard to introduce the carrying capacity $K$ of the environment inhabited by the population. We then get the logistic or Verhulst population model

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right), \quad (1.2)$$

which has the property as long as $N(0) \in (0, K]$ it holds that $N \in (0, K]$ for all $t > 0$.

Although these models describe the qualitative behaviour of many populations, it is erroneous to regard population growth as an entirely deterministic process. To remedy this, we would like to randomize the parameters of the models in an appropriate way. We can decompose the growth parameter in the exponential model as $r(t) = \bar{r}(t) + "\text{noise}"$ yielding

$$\frac{dN(t)}{dt} = \bar{r}(t) N(t) + N(t) \cdot "\text{noise}". \quad (1.3)$$

In differential form (1.3)) is written

$$dN(t) = \bar{r}(t) N(t) \, dt + N(t) \cdot "\text{noise}" \, dt,$$

which can be transformed into a stochastic differential equation

$$dN(t) = \bar{r}(t) N(t) \, dt + \sigma(t) N(t) \, dW(t), \quad (1.4)$$

where $W(t)$ represents a certain kind of normalized noise and $\sigma(t)$ the size of the noise in this particular case. The noise $W(t)$, known as the Wiener process or Brownian motion, consists of normally distributed random variables whose mean is zero and whose variance is equal to $t$. Note that the solution to (1.4)) is not a single function, but rather a sample of functions.
We shall not go further into the precise interpretation of (1.4)) here and will return to it later in the course when we have developed methods for finding solutions to stochastic differential equations.

Randomizing the Verhulst model poses a greater difficulty, at least if we randomize $r$ and $K$ independently. The carrying capacity must be positive, but if we decompose $K$ with a normally distributed random part we cannot guarantee that $K$ will remain positive. This illustrates that care must be taken when randomizing deterministic equations. Introducing $\tilde{N} = \frac{r}{K}N$ and $\lambda = r$, the Verhulst population model can be rescaled into

$$\frac{d\tilde{N}}{dt} = \lambda\tilde{N} - \tilde{N}^2.$$  

(1.5)

By randomizing $\lambda$ we can construct a stochastic Verhulst model as

$$d\tilde{N} (t) = \left( \lambda (t) \tilde{N} (t) - \left( \tilde{N} (t) \right)^2 \right) dt + \sigma (t) \tilde{N} (t) dW (t),$$  

(1.6)

where the noise is proportional to the size of the population just as in the case of the stochastic model of exponential growth. We shall return to (1.6) and its solution later in the course.

1.2 Pricing of stock options

A European call option gives the right, but not the obligation, to buy a share at a specified price $K$ at a fixed future time $T$. If we let $X (T)$ denote the value of the share at time $T$, the resulting payoff is given by

$$(X (T) - K)^+ = \begin{cases} X (T) - K, & \text{if } X (T) > K, \\ 0, & \text{if } X (T) \leq K. \end{cases}$$  

(1.7)

How much should a person be willing to pay for such an option?

Assume that the price of the share at some given time $t$ is known to be $x$. Due to the well-known Black-Scholes’ equation (see Example 5.75), the price $p (t, x)$ of a European call option is the solution to the partial differential equation with terminal value

$$\begin{cases} \frac{\partial p}{\partial t} + r (t) x \frac{\partial p}{\partial x} + \frac{(\sigma (t) x)^2}{2} \frac{\partial^2 p}{\partial x^2} = r (t) p, \\ p (T, x) = (x - K)^+, \end{cases}$$  

(1.8)

where $r (t)$ is the interest rate and $\sigma (t)$ is the volatility (size of random fluctuations) of the share. In some special cases this equation can be solved explicitly, but in most cases we must use numerical methods for partial differential equations, such as finite difference or finite element methods, to find the solution.

The problem of determining $p (t, x)$ can be approached from another direction as well. We can model the value $X (t)$ of the share by the stochastic differential equation

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

(1.9)

where $r (t)$ and $\sigma (t)$ are as above. By the Feynman-Kac formula (Theorem 5.13), the option price is given by the conditional expectation

$$p (t, x) = E \left[ e^{-r (T - t)} (X (T) - K)^+ | X (t) = x \right].$$  

(1.10)
If we generate a number of trajectories of the share price starting at $X(t) = x$ by some numerical discretization of (1.9), we can approximate the option price by taking the sample mean of $e^{-r(T-t)} (X(T) - K)^+$ over the set of trajectories.

This example illustrates the important connection between stochastic and partial differential equations allowing problems to be approached from two different angles, where (of course) the preferable angle varies from problem to problem.

### 1.3 Stochastic optimal control

Suppose that we have two investment possibilities:

1. A safe investment (e.g. a bond), where the price $Q(t)$ grows exponentially with time according to
   \[
   \frac{dQ}{dt} = \rho(t) Q, \tag{1.11}
   \]
   with $\rho(t) > 0$.

2. A risky investment (e.g. a share), where the price $S(t)$ evolves according to the stochastic differential equation
   \[
   dS(t) = r(t) S(t) \, dt + \sigma(t) S(t) \, dW(t), \tag{1.12}
   \]
   with $r(t) > \rho(t)$.

At each time instant we can choose how large a fraction $\alpha(t)$ of our wealth $X(t)$ that should be placed in the risky asset, thereby placing $(1-\alpha(t))X(t)$ in the safe asset. We would like to determine an optimal investment strategy $f(t)$ from $0 \leq t \leq T$ that maximizes our wealth at a given final time $T$. For some function $g$ (in the simplest case, $g(x) = x$), we would like to determine

\[
   u(t,x) = \max_\alpha E[g(X(T)) | X(t) = x]. \tag{1.13}
\]

In Section 6, we show that the solution to this so-called optimal control problem can be obtained by means of a Hamilton Jacobi equation, which in general is a partial differential equations of the form

\[
   \frac{\partial u}{\partial t} + H \left( u, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2} \right) = 0, \tag{1.14}
\]

where the Hamiltonian $H$ is some function, that can be derived from the setup of the optimal control problem. Again this is an example of the connection between stochastic and partial differential equations.

### 1.4 Probability and measure theory

We can take the tossing of a fair die as an example of a naive approach to probability theory. There are 6 possible outcomes which we denote by $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$ and $\omega_6$. We let the set of all possible outcomes \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\} constitute the sample space $\Omega$. If we toss the die a large number of times $N$, we expect the relative frequency of a certain outcome $\omega_i$, defined as $f_i(N) = N_i/N$, where $N_i$ is the number of occurrences of outcome $\omega_i$, to converge to the probability $p_i$ of outcome $\omega_i$, that is

\[
   \lim_{N \to \infty} f_i(N) = p_i, \tag{1.15}
\]
for \( i = 1, 2, ..., 6 \).

By an event we mean a subset of the sample space, for example \( \{ \omega_1 \} \) or \( \{ \omega_1, \omega_3, \omega_5 \} \), such that we can distinguish if the event has occurred or not. If a set \( A \) is an event, naturally so is \( A^c \). Furthermore if \( A \) and \( B \) are events so are \( A \cap B \) and \( A \cup B \) as well. In addition the whole sample space \( \Omega \) must be an event.

Observe that there is some flexibility in the choice of sample space. If we only register if the outcome of the die is odd or even, the sample space consists of \( \{ O, E \} \), where \( O = \{ \omega_1, \omega_3, \omega_5 \} \) and \( E = \{ \omega_2, \omega_4, \omega_6 \} \). In this case the set of events are \( \{ \emptyset, O, E, \Omega \} \), so \( \{ \omega_1 \} \) is not an event with this choice of sample space. We can think of the set of events as the amount of information that is available.

A natural generalization of the setup above is to consider a case with countably infinite sample space, for example the number of ants in an ant-hill at a given time. Although this number of course is finite it is nevertheless unbounded so, in principle, it can be infinite. Now the sample space is \( \{ \omega_1, \omega_2, ... \} \). If we define \( E_n = \{ \omega_2, \omega_4, ... \omega_{2n} \} \), then the countable sum

\[
\bigcup_{n=1}^{\infty} E_n,
\]

representing an even number of ants in the ant-hill, must also be an event.

It turns out that problems occur when constructing a theory for probabilities based on limits of relative frequencies (Bertrand’s paradox). In 1933 Kolmogorov, guided by the intuitive rules of probability discussed above, developed an axiomatic approach to probability theory. We will adopt the axiomatic approach for the remainder of this course.

A probability space is a triple \((\Omega, \mathcal{A}, P)\), where \( \Omega \) is a non-empty set (the sample space), \( \mathcal{A} \) is a \( \sigma \)-algebra (the set of events) and \( P : \mathcal{A} \to [0, 1] \) is a function that assigns probabilities to events (the probability measure).

**Definition 1.1** A collection \( \mathcal{A} \) of subsets of \( \Omega \) is a \( \sigma \)-algebra if

1. \( \Omega \in \mathcal{A} \).
2. \( A^c \in \mathcal{A} \), if \( A \in \mathcal{A} \).
3. \( \bigcup_{n=1}^{\infty} A_n \in \mathcal{A} \), if \( A_1, A_2, ... \in \mathcal{A} \).

A \( \sigma \)-algebra contains the whole space and is closed under the set operations of complementation and countable unions. It is easily shown that it contains the empty set and is closed under countable intersections as well.

For a given sample space \( \Omega \), there are many possible choices of the \( \sigma \)-algebra \( \mathcal{A} \). The smallest of these choices contains only \( \emptyset \) and \( \Omega \), whereas the largest contains all subsets of \( \Omega \). Any collection of subsets \( \mathcal{C} \) (\( \sigma \)-algebra or not) can be enlarged to a \( \sigma \)-algebra, by adding sets according to the three conditions in Definition 1.1. We call the resulting \( \sigma \)-algebra \( \mathcal{A}(\mathcal{C}) \) the \( \sigma \)-algebra generated by \( \mathcal{C} \). Note that this \( \sigma \)-algebra will be the smallest \( \sigma \)-algebra containing \( \mathcal{C} \). An important example of generated \( \sigma \)-algebras is the Borel \( \sigma \)-algebra \( \mathcal{B} \) generated by the collection of semi-infinite intervals \( (\infty, a] \) on \( \mathbb{R} \). Its elements are called Borel sets. All open sets, closed sets, countable unions or intersections of open and closed sets and so on are examples of Borel sets. The notion of Borel sets can be generalized to higher dimensions as well if we introduce product \( \sigma \)-algebras (the details of this procedure is left out).
We call the ordered pair \((\Omega, \mathcal{A})\) a measurable space. The reason for this notation is we can assign measures to measurable spaces. A measurable space \((\Omega, \mathcal{A})\) and a measure \(\mu\) defined on \((\Omega, \mathcal{A})\) together form a measure space \((\Omega, \mathcal{A}, \mu)\).

**Definition 1.2** A measure \(\mu\) on a measurable space \((\Omega, \mathcal{A})\) is a non-negative valued set function on \(\mathcal{A}\) satisfying

1. \(\mu(\emptyset) = 0\)

2. \(\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n),\) for any sequence \(A_1, A_2, \ldots \in \mathcal{A}\) with \(A_i \cap A_j = \emptyset\) for \(i \neq j\).

It follows that \(\mu(A) \leq \mu(B)\) for all \(A, B \in \mathcal{A}\) such that \(A \subseteq B\). A probability measure is a measure with the additional property \(\mu(\Omega) = 1\). If a measure \(\mu\) is finite, i.e., if \(\mu(\Omega) < \infty\), it can be used to construct a probability measure as \(P(A) = \mu(A)/\mu(\Omega)\). Conclusively a probability space is a measurable space equipped with a probability measure (or equivalently a measure space \((\Omega, \mathcal{A}, \mu)\) for which the measure is a probability measure).

The \(\sigma\)-algebra of a probability space may contain nonempty events \(A\) for which \(P(A) = 0\) and we call such events null events. The sample space is also referred to as the sure event and any other event \(A\) satisfying \(P(A) = 1\) is said to occur almost surely (a.s.) or with probability one (w.p.1).

From the definition of probability measures one can derive the Borel-Cantelli lemma

**Lemma 1.3** Let \((\Omega, \mathcal{A}, P)\) be a probability space and let \(\{A_n\}_{n=1}^{\infty}\) be an infinite sequence of events in \(\mathcal{A}\). Then

\[
\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 0. \tag{1.17}
\]

The set \(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\) represents the set of outcomes occurring in infinitely many of the events \(A_n\). The Borel-Cantelli lemma indicates that if \(P(A_n)\) converges, the outcomes occurring infinitely often form a set of measure zero.

A measure space \((\Omega, \mathcal{A}, \mu)\) is not complete in the sense that the definition does not rule out the existence of subsets \(A^*\) of \(\Omega\) such that \(A^* \notin \mathcal{A}\) but \(A^* \subset A \in \mathcal{A}\) and \(\mu(A) = 0\). By adding all sets in \(\Omega\) with measure zero to \(\mathcal{A}\) and then extending \(\mathcal{A}\) according to Definition 1.1, one can however always make \((\Omega, \mathcal{A}, \mu)\) complete. We shall assume that such an extension has been carried out on all measure spaces considered in this course.

An important measure is the Borel measure \(\mu_B\) defined on the Borel \(\sigma\)-algebra. To every finite interval \([a, b]\) it assigns the length \(\mu_B([a, b]) = b - a\). Completing the measure space \((\mathbb{R}, \mathcal{B}, \mu_B)\) according to the procedure above we get the \(\sigma\)-algebra of Lebesgue subsets \(\mathcal{L}\) and the Lebesgue measure \(\mu_L\). The Borel and Lebesgue measures can also be generalized to higher dimensions if we introduce product measures (the details of this procedure are left out).

Given two measurable spaces \((\Omega_1, \mathcal{A}_1)\) and \((\Omega_2, \mathcal{A}_2)\) respectively, we say that a function \(f : \Omega_1 \to \Omega_2\) is \(\mathcal{A}_1: \mathcal{A}_2\)-measurable if

\[
f^{-1}(A_2) = \{\omega \in \Omega_1 : f(\omega) \in A_2\} \in \mathcal{A}_1 \tag{1.18}
\]

for all \(A_2 \in \mathcal{A}_2\). In most cases of interest the measurable space \((\Omega_2, \mathcal{A}_2)\) will be either \((\mathbb{R}, \mathcal{B})\) or \((\mathbb{R}, \mathcal{L})\) and if so we say that a function is \(\mathcal{A}_1\)-measurable (sometimes Borel- or Lebesgue \(\mathcal{A}_1\)-measurable).
For any function $X : \Omega \to \mathbb{R}$, we can create the $\sigma$-algebra $\mathcal{A}(X)$ generated by $X$ as the smallest $\sigma$-algebra containing all the sets

$$X^{-1}(L),$$

for every $L \in \mathcal{L}$. By construction $X$ will be $\mathcal{A}(X)$-measurable.

**Definition 1.4** Let $(\Omega, \mathcal{A}, P)$ be a probability space. A random variable $X$ is an $\mathcal{A}$-measurable function $X : \Omega \to \mathbb{R}$.

Every random variable induces a probability measure on $(\mathbb{R}, \mathcal{L})$ by means of

$$P_X(L) = P(X^{-1}(L)),$$

for every $L \in \mathcal{L}$. This probability measure is known as the distribution of $X$ and it indicates the likelihood that the function value of $X$ can be found in the set $L$. Note that the distribution is connected to the well-known distribution function of $X$, $F_X(x)$, by the relation $F_X(x) = P_X((\infty, x])$.

**Definition 1.5** Two subsets $A, B \in \mathcal{A}$ are called independent if

$$P(A \cap B) = P(A)P(B).$$

A collection $\mathcal{H}$ of families of $\mathcal{A}$-measurable sets $\mathcal{H}_i$ are independent if

$$P(H_{i_1} \cap H_{i_2} \cap \ldots \cap H_{i_k}) = P(H_{i_1})P(H_{i_2})\cdots P(H_{i_k}),$$

for all choices $H_{i_1} \in \mathcal{H}_{i_1}, \ldots, H_{i_k} \in \mathcal{H}_{i_k}$ with different indices $i_1, \ldots, i_k \in \{1, 2, \ldots, k\}$. A collection of random variables $X_1, \ldots, X_k$ is independent if the collection of $\sigma$-algebras generated by the random variables $\{\mathcal{A}(X_1), \ldots, \mathcal{A}(X_k)\}$ is independent.

### 1.5 The Lebesgue Integral

The well-known Riemann integral is defined as the common limit of lower and upper rectangular approximations (see page 55 in [KP92] for a repetition). All continuous functions and all discontinuous functions with finitely many jumps are Riemann integrable.

Nevertheless many quite simple function are not Riemann integrable. One example is the indicator function $I_{\mathbb{Q}}$ of the set of rational numbers, defined as

$$I_{\mathbb{Q}}(x) = \begin{cases} 1, & \text{for } x \in \mathbb{Q} \\ 0, & \text{for } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

On any bounded interval $[a, b]$ the lower rectangular approximation of $I_{\mathbb{Q}}$ will be zero and the upper approximation will be $b - a$, so $I_{\mathbb{Q}}$ cannot be Riemann integrable. Note that $\mathbb{Q}$ is a countable union of point sets, which are closed, and hence $\mathbb{Q}$ is a Borel set. From this fact we deduce that $I_{\mathbb{Q}}$ is a Lebesgue measurable function. Measurable functions play a fundamental part in the axiomatic approach to probability and, as many concepts in probability theory such as expectations and moments ($E[X^n]$) are defined as integrals, it is desirable that we find a way to integrate all measurable functions. The solution to this problem is provided by the Lebesgue integral.
Let \((\Omega, \mathcal{A}, \mu)\) be a measure space and suppose that \(\mu\) is a finite. Proceeding step by step we shall define the Lebesgue integral for all real-valued \(\mathcal{A}\)-measurable functions on this space. A simple function on \((\Omega, \mathcal{A}, \mu)\) is a real valued function on the form
\[
\phi = \sum_{i=1}^{M} a_i I_{A_i},
\]  
for \(a_1, a_2, \ldots, a_M \in \mathbb{R}\) and mutually disjoint sets \(A_1, A_2, \ldots, A_M \in \mathcal{A}\). By construction \(\phi\) is \(\mathcal{A}\)-measurable. The Lebesgue integral of a simple function over a set \(A \in \mathcal{A}\) is given by
\[
\int_{A} \phi \, d\mu = \sum_{i=1}^{N} a_i \mu(A \cap A_i).
\]  

If \(f\) is \(\mathcal{A}\)-measurable, nonnegative and bounded, that is if there exists some finite \(N\) such that \(0 \leq f(x) < N\) for all \(\omega \in \Omega\), we define the Lebesgue integral as
\[
\int_{A} f \, d\mu = \sup_{\phi \leq f} \int_{A} \phi \, d\mu,
\]  
where the supremum is taken over all simple functions with \(\phi(\omega) \leq f(\omega)\) for all \(\omega \in A\). This definition makes sense as we, for any nonnegative bounded function, can find a sequence of simple functions \(\{\phi_n\}_{n=1}^{\infty}\) such that \(\phi_n(\omega) \to f(\omega)\) as \(n \to \infty\) for all \(\omega \in A\). One such sequence of simple functions is given on page 56 in [KP92].

If \(f\) is \(\mathcal{A}\)-measurable and nonnegative (but not necessarily bounded), we let
\[
\int_{A} f \, d\mu = \sup_{N} \int_{A} f_N \, d\mu,
\]  
where \(f_N(\omega) = \min(f(\omega), N)\) is a \(\mathcal{A}\)-measurable, nonnegative and bounded function, which can be integrated by means of (1.26). Note that the supremum in (1.27) can become infinite. Finally, for \(\mathcal{A}\)-measurable functions \(f\), we write \(f = f^+ - f^-\), where \(f^+(\omega) = \max(f(\omega), 0)\) and \(f^-(\omega) = \max(-f(\omega), 0)\) are \(\mathcal{A}\)-measurable and nonnegative functions, which can be integrated as in (1.27). Conclusively, the Lebesgue integral of a \(\mathcal{A}\)-measurable function \(f\) can be written
\[
\int_{A} f \, d\mu = \int_{A} f^+ \, d\mu - \int_{A} f^- \, d\mu.
\]  
The set of all Lebesgue integrable functions, that is \(\mathcal{A}\)-measurable functions for which the Lebesgue integral is finite, is denoted \(L^1(\Omega, \mathcal{A}, \mu)\).

The Riemann and Lebesgue integrals coincide for all functions that are Riemann integrable (see page 58 in [KP92] for a proof). Furthermore it follows from the linearity of the Lebesgue integral for simple functions that the basic properties of the Riemann integral carries over to the Lebesgue integral.

**Proposition 1.6** Let \(f, g : \Omega \to \mathbb{R}\) belong to \(L^1(\Omega, \mathcal{A}, \mu)\) and let \(\alpha, \beta \in \mathbb{R}\). Then \(\alpha f + \beta g \in L^1(\Omega, \mathcal{A}, \mu)\) and for any \(A \in \mathcal{A}\)

1. \(\int_{A} (\alpha f + \beta g) \, d\mu = \alpha \int_{A} f \, d\mu + \beta \int_{A} g \, d\mu,\)

2. \(\int_{A} f \, d\mu \leq \int_{A} g \, d\mu\) if \(f(\omega) \leq g(\omega)\) for almost all \(\omega \in A\).
3. \(|\int_A f \, d\mu| \leq \int_A |f| \, d\mu\)

By almost all \(\omega \in A\) we mean that the statement is true for all \(\omega \in A\) except for those belonging to a set \(B \subset A\) with \(\mu(B) = 0\). In addition, for any disjoint sets \(A, B \in \mathcal{A}\), we have

\[
\int_{A \cup B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu. \quad (1.29)
\]

A tool that will be used repeatedly is the Theorem of dominated convergence due to Lebesgue, which gives conditions under which limits and integrals can be interchanged.

**Theorem 1.7 (Dominated convergence, 2.2.3 in [KP92])** Let \(f, g : \Omega \to \mathbb{R}\) belong to \(L^1(\Omega, \mathcal{A}, \mu)\), where \(\mu(\Omega) < \infty\) and let \(\{f_n\}_{n=1}^\infty\) be a sequence of \(\mathcal{A}\)-measurable functions satisfying \(f_n(\omega) \leq g(\omega)\) and

\[
\lim_{n \to \infty} f_n(\omega) = f(\omega), \quad (1.30)
\]

for almost all \(\omega \in \Omega\). Then

\[
\lim_{n \to \infty} \int_\Omega f_n \, d\mu = \int_\Omega f \, d\mu. \quad (1.31)
\]

### 1.6 Expectation

Our first application of the Lebesgue integral will be to define the expectation (mean value) of a random variable.

**Definition 1.8** Let \(X : \Omega \to \mathbb{R}\) be a random variable on the probability space \((\Omega, \mathcal{A}, P)\) and let \(P_X\) denote its distribution on \((\mathbb{R}, \mathcal{L})\). For \(X \in L^1(\Omega, \mathcal{A}, P)\) we define the expectation of \(X\) as

\[
E[X] = \int_\Omega X(\omega) \, dP(\omega) = \int_\mathbb{R} x \, dP_X(x). \quad (1.32)
\]

Note that when then the distribution function \(F_X(x) = P_X((-\infty, x])\) is differentiable, the expectation as defined in (1.32) reduces to

\[
E[X] = \int_\mathbb{R} xp(x) \, dx, \quad (1.33)
\]

where \(p(x) = dF_X(x)/dx\) is the density function of \(X\). This is the familiar formula for the expectation of a random variable.

Since the expectation is defined as an integral it inherits all the basic properties of integrals found in Proposition 1.6 above. From these properties it is clear that we can integrate (and hence calculate the expectation) of sums of integrable random variables. What about more complicated functions of random variables? The Doob-Dynkin lemma (Proposition 3 on page 7 in [Rao84]) asserts that if \(X, Y : \Omega \to \mathbb{R}\) are two given functions, then \(Y\) is \(\mathcal{A}(X)\)-measurable if and only if \(Y\) can be written \(Y = g(X)\) for some Borel measurable function \(g : \mathbb{R} \to \mathbb{R}\). Thus for all random variables \(X\) and Borel measurable functions \(g : \mathbb{R} \to \mathbb{R}\) such that \(g(X) \in L^1(\Omega, \mathcal{A}, P)\) the expectation

\[
E[g(X)] = \int_\Omega g(X(\omega)) \, dP(\omega) = \int_\mathbb{R} g(x) \, dP_X(x), \quad (1.34)
\]

is well defined.

In order to be able to calculate higher order moments \(E[X^p]\) of random variables, such as the variance, we must ascertain that \(X^p \in L^1(\Omega, \mathcal{A}, P)\). For this purpose we introduce the \(L^p\)-spaces.
**Definition 1.9** Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable on the probability space $(\Omega, \mathcal{A}, P)$ and let $p \in [1, \infty)$. Then $X \in L^p(\Omega, \mathcal{A}, P)$ if

$$
\|X\|_p := \left( \int_{\Omega} |X(\omega)|^p \, dP(\omega) \right)^{\frac{1}{p}} = \left( E[|X|^p] \right)^{\frac{1}{p}} < \infty.
$$

(1.35)

The norm $\|X\|_p$ is known as the $L^p$-norm of $X$.

By Definition 1.9, it is clear that the moment $E[X^p]$ is well defined for all $X \in L^p(\Omega, \mathcal{A}, P)$. The $L^p$-spaces are complete in the sense that Cauchy sequences in a $L^p$-space converge to an element in that space. We conclude this Section by stating a number of results that will prove useful later on.

**Proposition 1.10** Let $X, Y \in L^1(\Omega, \mathcal{A}, P)$ be two independent random variables. Then

$$
E[XY] = E[X]E[Y].
$$

(1.36)

**Proposition 1.11** Let $X, Y \in L^2(\Omega, \mathcal{A}, P)$ be random variables. The Cauchy-Schwarz inequality then states that $XY \in L^1(\Omega, \mathcal{A}, P)$ and

$$
E[XY] \leq (E[X^2])^{\frac{1}{2}}(E[Y^2])^{\frac{1}{2}}.
$$

(1.37)

The Cauchy-Schwarz inequality is a special case of the more general Hölder inequality, which states the following.

**Proposition 1.12** Let $X \in L^p(\Omega, \mathcal{A}, P)$ and $Y \in L^q(\Omega, \mathcal{A}, P)$ be random variables, where $\frac{1}{p} + \frac{1}{q} = 1$. Then $XY \in L^1(\Omega, \mathcal{A}, P)$ and

$$
E[XY] \leq (E[X^p])^{\frac{1}{p}}(E[Y^q])^{\frac{1}{q}}.
$$

(1.38)

**Proposition 1.13** Let $X \in L^1(\Omega, \mathcal{A}, P)$ be a random variable and let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function. Jensen’s inequality then states that

$$
g(E[X]) \leq E[g(X)]
$$

(1.39)

The inequalities in Propositions 1.11-1.13 are direct consequences of similar inequalities for integrals of functions $f : \Omega \rightarrow \mathbb{R}$ (see page 22 in [KP92]). The following two inequalities give an upper bound of the probability for large deviations of a random variable.

**Proposition 1.14** Let $X \in L^1(\Omega, \mathcal{A}, P)$ be a random variable which is positive for almost all $\omega$. Then the Markov inequality states that for all $a > 0$

$$
P(\omega \in \Omega : X(\omega) \geq a) \leq \frac{1}{a}E[X].
$$

(1.40)

**Proposition 1.15** A similar result is the Chebyshev inequality, which states that if $X \in L^p(\Omega, \mathcal{A}, P)$ for some $p > 0$, then for all $a > 0$

$$
P(\omega \in \Omega : |X(\omega)| \geq a) \leq \frac{1}{a^p}E[|X|^p].
$$

(1.41)
1.7 Conditional expectation

In most cases we will use the Lebesgue measure \( \mu_L \) as the measure \( \mu \) in the Lebesgue integral. We can, however, use any other measure and of particular interest are measures \( \mu \) that are absolutely continuous with respect to \( \mu_L \), denoted \( \mu \ll \mu_L \). By absolutely continuous we mean that \( \mu (L) = 0 \) for all sets \( L \in \mathcal{L} \) such that \( \mu_L (L) = 0 \).

Any measurable function \( p : \mathbb{R} \rightarrow \mathbb{R} \) being non-negative for almost all \( \omega \in \Omega \) can be used to construct a measure which is absolutely continuous with respect to \( \mu_L \) by means of

\[
\mu (A) = \int_A p \, d\mu_L, \tag{1.42}
\]

for all \( A \in \mathcal{L} \). The converse of this statement is known as the Radon-Nikodym Theorem. For every measure \( \mu \) being absolutely continuous with respect to the Lebesgue measure, it asserts the existence of a unique (for almost all \( \omega \in \Omega \)) measurable function \( p \) satisfying (1.42). The function \( p \) is known as the Radon-Nikodym derivative because formally we can write \( p = \frac{d\mu}{d\mu_L} \).

The Radon-Nikodym derivative is of interest here because it gives us a way to define conditional expectations in axiomatic probability theory.

**Definition 1.16** Let \( X : \Omega \rightarrow \mathbb{R} \) be an integrable random variable on the probability space \( (\Omega, \mathcal{A}, P) \). Furthermore let \( S \subset \mathcal{A} \) be a \( \sigma \)-algebra (containing less information than \( \mathcal{A} \)). Then the conditional expectation of \( X \) with respect to \( S \), denoted \( E [X|S] \), is the unique (for almost all \( \omega \in \Omega \)) \( S \)-measurable function satisfying

\[
\int_S X \, dP = \int_S E [X|S] \, dP, \tag{1.43}
\]

for all \( S \in \mathcal{S} \).

If \( X \) is positive for almost all \( \omega \in \Omega \), we can define a measure on \( S \) as

\[
\mu (S) = \int_S X \, dP, \tag{1.44}
\]

for all \( S \in \mathcal{S} \) (if \( X \) is not positive we can make the same conclusions for \( X^+ \) and \( X^- \)). Clearly \( \mu \ll P \) on \( S \), so the Radon-Nikodym theorem asserts the existence of a Radon-Nikodym derivative, that is a \( S \)-measurable function \( p \) satisfying

\[
\int_S X \, dP = \mu (S) = \int_S p \, dP, \tag{1.45}
\]

Setting \( E [X|S] = p \) the definition of conditional expectation makes sense.

The conditional expectation can be seen as the best approximation of \( X \) that can be achieved with the incomplete information supplied by \( S \). We have the following basic results for conditional expectations.

**Proposition 1.17** Let \( X, Y : \Omega \rightarrow \mathbb{R} \) be integrable random variables on the probability space \( (\Omega, \mathcal{A}, P) \) and let \( S \subset \mathcal{A} \) be a \( \sigma \)-algebra. For \( \alpha, \beta \in \mathbb{R} \), we have

1. \( E [aX + bY|S] = aE [X|S] + bE [Y|S] \)
2. \( E [E [X|S]] = E [X] \),
3. \( E[X|S] = X \), if \( X \) is \( S \)-measurable,

4. \( E[X|S] = E[X] \), if \( X \) is independent of \( S \),

5. \( E[XY|S] = YE[X|S] \), if \( Y \) is \( S \)-measurable.

If \( T \) is another \( \sigma \)-algebra such that \( S \subset T \subset A \), we have in addition

\[
E[X|S] = E[E[X|T]|S], \tag{1.46}
\]

Most basic properties of the expectation are carried over to conditional expectations. It is also possible to condition on random variables instead of on \( \sigma \)-algebras. If \( X, Y \) are random variables on \((\Omega, \mathcal{A}, P)\) we define the expectation of \( X \) conditioned on \( Y \), written \( E[X|Y] \), as \( E[X|\mathcal{A}(Y)] \), where \( \mathcal{A}(Y) \) is the \( \sigma \)-algebra generated by \( Y \).

### 1.8 Convergence of random sequences

In many of the proofs carried out in this course we will study the convergence of sequences \( \{X_n\}_{n=1}^{\infty} \) of random variables defined on a common probability space \((\Omega, \mathcal{A}, P)\). The most commonly used modes of convergence are

1. **Convergence with probability one (w.p.1)** (also known as almost sure convergence):

   \[
P \left( \left\{ \omega \in \Omega : \lim_{n \to \infty} |X_n(\omega) - X(\omega)| = 0 \right\} \right) = 1. \tag{1.47}
\]

2. **Mean-square convergence**: \( E[X^2] < \infty, E[X_n^2] < \infty \) for \( n = 1, 2, \ldots \) and

   \[
   \lim_{n \to \infty} E[|X_n - X|^2] = 0. \tag{1.48}
   \]

3. **Convergence in probability**:

   \[
   \lim_{n \to \infty} P(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0, \tag{1.49}
   \]

   for all \( \varepsilon > 0 \).

The convergence modes 1 and 2 both imply 3, but apart from that nothing general can be said about the pairwise relations between the convergence modes. However, if \( |X_n(\omega)| \leq Y(\omega) \) for almost all \( \omega \in \Omega \), where \( Y \) is some random variable satisfying \( E[Y^2] < \infty \), then the Theorem of dominated convergence can be used to show that convergence modes 2 and 3 are equivalent and, as a consequence, that 1 implies 2.

### 2 Stochastic processes

We begin with a definition.

**Definition 2.1** A stochastic process \( X = \{X_t : t \in T\} \), with \( X_t : \Omega \to \mathbb{R} \), is a collection of random variables on a common probability space \((\Omega, \mathcal{A}, P)\) indexed by a parameter \( t \in T \subset \mathbb{R} \).
When $T$ is a finite or countable set, the stochastic process is just a sequence of random variables $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$, which can be thought of as the evolution of a probabilistic system over the discrete instants of time $t_1 < t_2 < \ldots < t_n$. The theory of such stochastic processes in discrete time is developed in Section 1.4 in [KP92] and the behaviour of discrete time stochastic processes depends on the joint distribution of all the random variables in the process.

For stochastic processes in continuous time, that is when $T$ is an uncountable set, for example an interval, the situation becomes more complicated. In this case no joint distribution with a finite number of variables can fully describe the behaviour at all time instants of the uncountable set $T$. This fact makes it difficult to state the relationship between $X_t$ at different time instants.

The following theory for continuous time stochastic processes is due to Kolmogorov. For each fixed time $t \in T$, the stochastic process $X$ defines a random variable

$$
\omega \rightarrow X_t(\omega),
$$

(2.1)

for all $\omega \in \Omega$. On the other hand, fixing $\omega \in \Omega$, the function

$$
t \rightarrow X_t(\omega),
$$

(2.2)

defined for all $t \in T$ constitutes a sample path (or trajectory or realization) of the stochastic process. With this image, $X_t(\omega)$ represents the position of sample path $\omega$ at time $t$. Each $\omega$ then represents a function from $T$ to $\mathbb{R}$, so we can regard $\Omega$ as a subset of the space of all such functions (which we denote $\mathbb{R}^T$). The $\sigma$-algebra $\mathcal{A}$ then contains the $\sigma$-algebra $\mathcal{A}'$ generated by the sets

$$
\{ \omega \in \Omega : X_{t_1}(\omega) \in L_1, \ldots, X_{t_n}(\omega) \in L_n \},
$$

(2.3)

for any $t_i \in T$ and $L_i \in \mathcal{L}$. In this framework the stochastic process can be viewed as a probability measure on the measurable space $(\mathbb{R}^T, \mathcal{A}')$.

The finite joint distribution associated with this probability measure is

$$
P_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(L_1 \times L_2 \times \cdots \times L_n) = P(X_{t_1}(\omega) \in L_1, \ldots, X_{t_n}(\omega) \in L_n),
$$

(2.4)

for $t_i \in T$ and $L_i \in \mathcal{L}$ and the finite joint distribution function is

$$
F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_1, x_2, \ldots, x_n) = P(X_{t_1}(\omega) \in (-\infty, x_1], \ldots, X_{t_n}(\omega) \in (-\infty, x_n]),
$$

(2.5)

for $t_i \in T$ and $x_i \in \mathbb{R}$.

The joint distribution determines many (but not all) properties of a stochastic process. It is also interesting to know when this procedure can be reversed, that is when a stochastic process can be created from a given joint distribution. Kolmogorov’s extension theorem gives the answer.

**Theorem 2.2** For all $t_1, \ldots, t_k \in T$, $k \in \mathbb{N}$ let $Q_{t_1, \ldots, t_k}$ be probability measures on $\mathbb{R}^T$ that satisfies the symmetry property

$$
Q_{t_{\sigma(1)}, \ldots, t_{\sigma(k)}}(L_1 \times \cdots \times L_k) = Q_{t_1, \ldots, t_k}(L_{\sigma^{-1}(1)} \times \cdots \times L_{\sigma^{-1}(k)}),
$$

(2.6)

for any permutation $\sigma$ of $\{1, 2, \ldots, k\}$, and the compatibility property

$$
Q_{t_1, \ldots, t_k, t_{k+1}, \ldots, t_{k+m}}(L_1 \times \cdots \times L_k \times \mathbb{R} \times \cdots \times \mathbb{R}) = Q_{t_1, \ldots, t_k}(L_1 \times \cdots \times L_k),
$$

(2.7)

for all $m \in \mathbb{N}$.

Then there exists a probability space $(\Omega, \mathcal{A}, P)$ and a stochastic process $X = \{X_t : t \in T\}$ with $Q_{t_1, \ldots, t_k}$ as its joint distribution.
The use of the finite joint distribution to characterize stochastic processes suggests the following definition of equality between two stochastic processes.

**Definition 2.3** Let \( X = \{ X_t : t \in T \} \), \( Y = \{ Y_t : t \in T \} \) be two stochastic processes on a common probability space \(( \Omega, A, P )\). \( X \) and \( Y \) are equivalent if

\[
P ( \{ \omega \in \Omega : X_t ( \omega ) = Y_t ( \omega ) \} ) = 1,
\]

for all \( t \in T \). In this case we say that \( X \) is a version of \( Y \) and vice versa.

Two equivalent stochastic processes have the same joint distribution (because the set in which they differ has measure zero). Nevertheless the sample paths of two versions of a process do not have to coincide.

In analogue to the three convergence modes of sequences of random variables discussed in Section 1.8, one can define three continuity modes of a continuous time stochastic process \( X \). If for any sequence \( \{ h_i \} \) of real numbers converging to zero, the sequence \( \{ X_{t+h_i} \}_{i=1}^{\infty} \) converges to \( X_t \) with probability one, we say that \( X \) is continuous with probability one. Similar definitions hold for the other two convergence modes. The extension of these continuity modes from pointwise to uniform continuity are carried out in exactly the same manner as for ordinary functions.

### 2.1 Markov and diffusion processes

A **Markov process** is a process where the future state, given the present state, is independent of the past. Hence, a discrete time Markov process \( \{ X_{t_n} \}_{n=1}^{\infty} \), where \( 0 < t_1 < t_2 < \ldots < t_n < t_{n+1} < \ldots \), is characterized by the (Markov) property

\[
P \left( X_{t_{n+1}} \in L | X_{t_n} = x_n, \ldots, X_{t_1} = x_1 \right) = P \left( X_{t_{n+1}} \in L | X_{t_n} = x_n \right).
\]

The future value \( X_{t_{n+1}} \) of the discrete time Markov process is given by the present state \( X_{t_n} \) and the transition probabilities

\[
P ( t_n, x_n : t_{n+1}, L ) = P \left( X_{t_{n+1}} \in L | X_{t_n} = x_n \right).
\]

Analogously the future value \( X_t \) of a continuous time Markov process is given by the present state \( X_s \) for some \( s < t \) and the transition probability

\[
P ( s, x : t, L ) = P \left( X_t \in L | X_s = x \right).
\]

When the continuous time Markov process satisfies sufficient smoothness properties, the role of transition probabilities can be replaced by **transition densities** \( p ( s, x : t, y ) \) defined as

\[
P ( s, x : t, L ) = \int_L p ( s, x : t, y ) \ dy.
\]

An important class of continuous time Markov processes, for which the transition probability satisfies certain regularity assumptions, is given by the diffusion processes.

**Definition 2.4** A continuous time Markov process is called a diffusion process if there exists functions \( a \) and \( b \), known as the drift and diffusion coefficients respectively, such that the transition probability satisfies the following three conditions for every \( s \in T \), \( x \in \mathbb{R} \) and \( \varepsilon > 0 \)
1. \( \lim_{t \to s} \frac{1}{t-s} \int_{|y-x|>\varepsilon} P(s, x : t, dy) = 0, \)

2. \( \lim_{t \to s} \frac{1}{t-s} \int_{|y-x|<\varepsilon} (y-x) P(s, x : t, dy) = a(s, x), \)

3. \( \lim_{t \to s} \frac{1}{t-s} \int_{|y-x|<\varepsilon} (y-x)^2 P(s, x : t, dy) = b^2(s, x). \)

The first condition asserts that the diffusion process is continuous w.p.1. The second and third conditions can be rewritten as

\[ a(s, x) = \lim_{t \to s} \frac{1}{t-s} E [X_t - X_s | X_s = x], \tag{2.13} \]

and

\[ b^2(s, x) = \lim_{t \to s} \frac{1}{t-s} E [(X_t - X_s)^2 | X_s = x], \tag{2.14} \]

respectively, showing that the drift and diffusion coefficients represent the rate of change of the mean and variance respectively per unit time. In the context of this course, diffusion processes are important due to the fact that all solutions to stochastic differential equations can be shown to be diffusion processes. We shall now examine the prototype of diffusion processes, the Brownian motion, more thoroughly.

### 2.2 Brownian motion

Brownian motion was first described by the botanist Robert Brown in 1828 as he studied pollen particles floating in water under the microscope. He observed minute particles within the vacuoles of the pollen grains executing a jittery motion. The physical explanation for the motion is, as Einstein pointed out in 1905, random collisions with water molecules. Mathematically, Brownian motion can be viewed as a stochastic process, known as the Wiener process.

Before defining the Wiener process we shall consider a number of important classes of stochastic processes.

**Definition 2.5** A stochastic process \( X = \{X_t : t \in T\} \) is said to have independent increments if the random variables

\[ X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_{n+1}} - X_{t_n}, \tag{2.15} \]

are independent for all choices of time instants \( t_0 < t_1 < \ldots < t_n \) in \( T \).

**Definition 2.6** A stochastic process \( X = \{X_t : t \in T\} \) is said to be Gaussian if its finite joint distribution is Gaussian (or normal), that is if there, for any finite subset \( \{t_1, \ldots, t_n\} \) of \( T \), exists a symmetric non-negative definite matrix \( C = [c_{jk}] \in \mathbb{R}^{n \times n} \) and a vector \( m \in \mathbb{R}^n \), such that the joint distribution function \( F_{X_{t_1}, \ldots, X_{t_n}} \) can be written

\[ F_{X_{t_1}, \ldots, X_{t_n}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(u_1, \ldots, u_n) \, du_n \cdots du_1, \tag{2.16} \]

with the density

\[ p(u_1, \ldots, u_n) = \frac{\sqrt{\det C}}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \sum_{j,k=1}^{n} (u_j - m_j) c_{jk} (u_k - m_k) \right). \tag{2.17} \]
Note that the vector \( \mathbf{m} \) represents the expectation and the matrix \( C^{-1} \) the covariance of the process.

**Definition 2.7** A stochastic process \( X = \{X_t : t \in T\} \) is said to be stationary if its joint distributions are invariant under time displacements, that is if

\[
P_{t_1+h, \ldots, t_n+h} = P_{t_1, \ldots, t_n},
\]

for all \( t_i, t_i + h \) in \( T \), where \( i \in \{1, 2, \ldots, n\} \) for some \( n \in \mathbb{N} \).

There are several equivalent definitions of the Wiener process and we shall choose a definition based on the properties of the process.

**Definition 2.8** The standard Wiener process \( W : [0, 1) \to \mathbb{R} \) is a Gaussian stochastic process with independent increments for which

1. \( W_0 = 0 \) w.p.1,
2. \( \mathbb{E}[W_t - W_s] = 0 \),
3. \( \mathbb{E}[(W_t - W_s)^2] = t - s \).

Note that the increments of the Wiener process are stationary. The transition density of the Wiener process is

\[
p(s, x : t, y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp \left( -\frac{(y-x)^2}{2(t-s)} \right),
\]

for all \( t > s \). One can generalize this definition and define a \( m \)-dimensional Wiener process \( (W^{(1)}, \ldots, W^{(m)}) \) as a collection of \( m \) independent one-dimensional Wiener processes. The transition density in this case is

\[
p(s, x : t, y) = (2\pi(t-s))^{-\frac{m}{2}} \exp \left( -\frac{|y-x|^2}{2(t-s)} \right),
\]

for all \( t > s \). It is shown in Section 1.8 in [KP92] that the Wiener process can be constructed as the limit of a random walk. This approach gives an intuitive image of the sample paths of the Wiener process. But will the sample paths be continuous and/or differentiable? The question of continuity is settled by the following lemma due to Kolmogorov.

**Lemma 2.9** A continuous time stochastic process \( X : [0, \infty) \times \Omega \to \mathbb{R} \) has a continuous version if there exist positive constants \( \alpha, \beta, C \) and \( h \) such that

\[
\mathbb{E}[|X_t - X_s|^\alpha] \leq C |t-s|^{1+\beta},
\]

for all \( s, t \in [0, \infty) \) with \( |t-s| \leq h \).

For the Wiener process, it follows from the transition density that

\[
\mathbb{E}[|W_t - W_s|^4] = 3|t-s|^2,
\]
so the prerequisites of the lemma are satisfied and there exist a continuous version of \( W \). From now on we suppose that \( W \) is this continuous version.

Although the paths of the Wiener process are continuous, they are nowhere differentiable (see Section 1.8 in [KP92] for a proof) and of unbounded variation on any finite interval. To explain the last statement, we consider arbitrary partitions \( a = x_0^{(n)} < x_1^{(n)} < \ldots < x_n^{(n)} = b \) of the interval \([a, b]\) where \( \delta^{(n)} = \max_{1 \leq i \leq n} \left(x_i^{(n)} - x_{i-1}^{(n)}\right) \to 0 \) as \( n \to \infty \). The total variation \( V_a^b (W) \) of \( W \) satisfies

\[
V_a^b (W) = \lim_{n \to \infty} \sum_{i=0}^{n-1} \left| W_{x_{i+1}^{(n)}} - W_{x_i^{(n)}} \right| = \infty,
\]  
and a function with infinite total variation is said to be of unbounded variation. In other words, the Wiener process varies infinitely on any finite interval, regardless of the size of the interval.

Applying the properties of Definition 2.8 to (2.13)-(2.14), it follows that the drift and diffusion coefficients of the Wiener process are zero and one respectively. Furthermore the sample paths of the Wiener process are continuous so it must be a diffusion process.

### 3 Stochastic integrals

#### 3.1 Background

In 1908 Langevin proposed that the position \( X_t \) at time \( t \) of a Brownian motion (or a similar process) could be described as a differential equation with a random term

\[
\frac{dX_t}{dt} = a(t, X_t) + b(t, X_t) \xi_t.
\]  

Here \( \xi_t \) is a Gaussian random variable (for reasons explained in Section 1.8 of [KP92], \( \xi_t \) is known as white noise) and \( a \) and \( b \) are functions, which should be related to the drift and diffusion coefficients of a diffusion process.

The differential equation above can be rewritten in integral form as

\[
X_t(\omega) = X_0 + \int_0^t a(s, X_s(\omega)) \, ds + \int_0^t b(s, X_s(\omega)) \xi_s(\omega) \, ds,
\]  
for each sample path \( \omega \). In Section 2.2, we saw that the drift and diffusion coefficients of the Wiener process are zero and one respectively, so if we set \( a \equiv 0 \) and \( b \equiv 1 \) in (3.2) we want the resulting process \( X \) to be the Wiener process. This implies that \( \xi \) must be the derivative of the Wiener process. But as we have seen the Wiener process is nowhere differentiable and hence \( \xi \) does not exist as an ordinary function. Formally we can write

\[
X_t(\omega) = X_0 + \int_0^t a(s, X_s(\omega)) \, ds + \int_0^t b(s, X_s(\omega)) \, dW_s(\omega),
\]  
but the rightmost integral is not a Lebesgue integral, since the continuous sample paths of the Wiener process are of unbounded variation (this makes it impossible to use \( W \) to create a measure used for integration). We need a new kind of integrals to solve this problem.

A rigorous definition of stochastic integrals

\[
\int_0^T f(s, \omega) \, dW_s(\omega),
\]  
18
was set up by Itô in 1951. As for the Lebesgue integral, the basic idea is to begin defining the stochastic integral for quite simple integrands and by various approximations work our way up to a wide class of integrands.

Before starting, however, there is an additional problem that must be addressed. Suppose that we would like to calculate the integral

\[ I = \int_0^T W_s(\omega) \, dW_s(\omega). \]  

(3.5)

We partition the interval of integration as

\[ 0 = t_0 < t_1 < \ldots < t_N = T \]

and test two different approximations of the integral \( I \), namely

\[ I_1 = \sum_{n=0}^{N-1} W_{t_n} (W_{t_{n+1}} - W_{t_n}) \quad \text{and} \quad I_2 = \sum_{n=0}^{N-1} W_{t_{n+1}} (W_{t_{n+1}} - W_{t_n}). \]  

(3.6)

If \( I \) would make sense as a Riemann integral - for example if \( W \) were differentiable - \( I_1 \) and \( I_2 \) would converge to the same integral, since the evaluation point is immaterial for the value of the Riemann integral. But in this stochastic case we get, due to the properties in Definition 2.8,

\[ E[I_1] = \sum_{n=0}^{N-1} E\left[W_{t_n} (W_{t_{n+1}} - W_{t_n})\right] = \sum_{n=0}^{N-1} E[W_{t_n}]E\left[W_{t_{n+1}} - W_{t_n}\right] = 0, \]

(3.7)

and

\[ E[I_2] = \sum_{n=0}^{N-1} E\left[W_{t_{n+1}} (W_{t_{n+1}} - W_{t_n})\right] \]

\[ = \sum_{n=0}^{N-1} \left( E\left[(W_{t_{n+1}} - W_{t_n})^2\right] + E[W_{t_n} (W_{t_{n+1}} - W_{t_n})] \right) = \sum_{n=0}^{N-1} (t_{n+1} - t_n) = T. \]  

The two approximations give completely different results, regardless of how fine we make the partition. Thus in order to get any consistency at all we must decide where the function is to be evaluated. This is a matter of taste, as different choices have different benefits. Choosing the left end point gives rise to the \textit{Itô integral}, which we shall define shortly. A stochastic integral where we, as the function evaluation, use the mean of the function evaluated at the left and right end points is known as a \textit{Stratonovich integral}, further discussed in Section 3.6.

### 3.2 The Itô integral

In order to state the class of integrands for which the Itô integral is defined, we need some preliminary definitions. The main restriction on the integrands \( f(t, \omega) \) is that they should depend only on the behaviour of \( W_s(\omega) \) for \( s \leq t \), that is only on the past and present behaviour of the Wiener process. We shall state this criterion more precisely in the following definitions.

The random variables \( W_t \) induced by the Wiener process \( W \) are \( \mathcal{A} : \mathcal{L}\)-measurable for each \( t \in T \), but most often the \( \sigma \)-algebra \( \mathcal{A} \) contain many more events than those detectable by \( W_t \) at a particular time instant, so it is interesting to consider the following more restricted \( \sigma \)-algebra as well.
Definition 3.1 Let $W$ be a Wiener process. Define the $\sigma$-algebra $A_t$ as the $\sigma$-algebra generated by the random variables $\{W_s\}_{0 \leq s \leq t}$. Hence $A_t$ will be the smallest $\sigma$-algebra containing all sets of the form

$$\{\omega \in \Omega : W_{t_1} \in L_1, \ldots, W_{t_k} \in L_k\}, \quad (3.9)$$

where $t_j \leq t$ and $L_j \in \mathcal{L}$.

One can think of $A_t$ as the history of $W$ up to time $t$. As time increases more information about $W$ is available, so naturally $A_s \subseteq A_t$ for $s < t$ and $A_t \subseteq A$ for all $t$. An increasing family of $\sigma$-algebras, such as $\{A_t\}$, is called a filtration.

Observe that since $W_t - W_s$ is independent of the history of the Wiener process up to time $s$, it holds that

$$E[W_t - W_s | A_s] = E[W_t - W_s] = 0, \quad (3.10)$$

or equivalently that

$$E[W_t | A_s] = E[W_s | A_s] = W_s. \quad (3.11)$$

Definition 3.2 Let $\{M_t\}_{t \geq 0}$ be a filtration. A stochastic process $g : [0, T] \times \Omega \rightarrow \mathbb{R}$ is said to be $M_t$-adapted if, for each $t \geq 0$, the function

$$\omega \rightarrow g(t, \omega), \quad (3.12)$$

is $M_t$-measurable.

In other words we say that a function is adapted to a filtration if the information in the filtration is enough to completely determine the value of the function. Consequently $h_t^{(1)}(\omega) = W_{t/2}(\omega)$ is $A_t$-adapted, while $h_t^{(2)}(\omega) = W_{2t}(\omega)$ is not.

Definition 3.3 For $0 < T < \infty$ we define the class $L^2_T$ of functions $f : [0, T] \times \Omega \rightarrow \mathbb{R}$ satisfying

1. $f$ is jointly $\mathcal{L} \times A$-measurable,
2. $f$ is $A_t$-adapted,
3. $\int_0^T E[(f(t, \cdot))^2] \, dt < \infty$,

Note that condition 3 in the above definition implies that $E[(f(t, \cdot))^2] < \infty$ for almost all $t \in [0, T]$. If we identify functions which differ only on a set on measure zero, the space $L^2_T$ is complete with the norm

$$\|f\|_{2,T} = \sqrt{\int_0^T E[(f(t, \cdot))^2] \, dt}. \quad (3.13)$$

Let $S^2_T$ denote the subset of all step functions in $L^2_T$, that is functions that can be written on the form

$$f(t, \omega) = f_j(\omega), \quad \text{for } t_j \leq t < t_{j+1}, \quad (3.14)$$

for some partition $0 = t_0 < t_1 < \ldots < t_n = T$. We define the Itô integral of step functions $f \in S^2_T$ as

$$I[f](\omega) = \int_0^T f(s, \omega) \, dW_s(\omega) = \sum_{j=0}^{n-1} f_j(\omega) (W_{t_{j+1}}(\omega) - W_{t_j}(\omega)). \quad (3.15)$$
Statement three will be referred to as the Itô isometry. The measurability of the Itô integral follows since it also belongs to \( S_T^2 \), whereas \( W_{t_j} \) is not \( \mathcal{A}_T \)-measurable and thus cannot be Itô integrable.

The Itô integral for \( f \in S_T^2 \) has the following properties.

**Lemma 3.4 (3.2.2 in [KP92])** For any \( f, g \in S_T^2 \) and \( \alpha, \beta \in \mathbb{R} \), the Itô integral satisfies

1. \( I[f] \) is \( \mathcal{A}_T \)-measurable,
2. \( E[I[f]] = 0 \),
3. \( E[(I[f])^2] = \int_0^T E[(f(t, \cdot))^2] \, dt \),
4. \( I[\alpha f + \beta g] = \alpha I[f] + \beta I[g] \), w.p.1.

Note that statement two and three provide formulas for the mean and variance of Itô integrals. Statement three will be referred to as the Itô isometry.

**Proof.** The measurability of the Itô integral follows since \( f_j \) is \( \mathcal{A}_T \)-measurable, \( W_{t_{j+1}} - W_{t_j} \) is \( \mathcal{A}_{t_{j+1}} \)-measurable and \( \mathcal{A}_i \subset \mathcal{A}_{t_{j+1}} \subset \mathcal{A}_T \). We also have

\[
E[I[f]] = E \left[ \sum_{j=0}^{n-1} f_j (W_{t_{j+1}} - W_{t_j}) \right] = \sum_{j=0}^{n-1} E \left[ f_j (W_{t_{j+1}} - W_{t_j}) | \mathcal{A}_j \right] \\
= \sum_{j=0}^{n-1} E \left[ f_j E \left[ (W_{t_{j+1}} - W_{t_j}) | \mathcal{A}_j \right] \right] = 0, \tag{3.16}
\]

by the properties of conditional expectations and the Wiener process. In the same manner

\[
E[(I[f])^2] = E \left[ \left( \sum_{j=0}^{n-1} f_j (W_{t_{j+1}} - W_{t_j}) \right)^2 \right] = \sum_{j=0}^{n-1} E \left[ f_j^2 (W_{t_{j+1}} - W_{t_j})^2 \right] = A_1 \\
+ \sum_{i<j} 2 E \left[ f_i f_j (W_{t_{i+1}} - W_{t_i}) (W_{t_{j+1}} - W_{t_j}) \right], \tag{3.17}
\]

where the two parts can be calculated as

\[
\sum_{j=0}^{n-1} A_1 = \sum_{j=0}^{n-1} E \left[ f_j^2 E \left[ (W_{t_{j+1}} - W_{t_j})^2 | \mathcal{A}_j \right] \right] = \sum_{j=0}^{n-1} E \left[ f_j^2 (t_{j+1} - t_j) \right] = \int_0^T E \left[ (f(t, \cdot))^2 \right] \, dt, \tag{3.18}
\]

and

\[
\sum_{i<j} A_2 = 2 \sum_{i<j} E \left[ f_i f_j (W_{t_{i+1}} - W_{t_i}) E \left[ (W_{t_{j+1}} - W_{t_j}) | \mathcal{A}_j \right] \right] = 0, \tag{3.19}
\]

respectively. This proves the Itô isometry. The fourth statement follows by the linearity of the definition of the Itô integral. \[Q.E.D.\]
We shall show that the Itô integral can be defined on $L_T^2$ as the limit of sequences of Itô integrals in $S_T^2$. As a first step we show that the space $S_T^2$ is dense in $L_T^2$, that is any function in $L_T^2$ can be approximated arbitrarily close, in terms of the norm in $L_T^2$, by functions in $S_T^2$.

Lemma 3.5 (3.2.1 in [KP92]) The space $S_T^2$ is dense in $L_T^2$.

Proof. First we approximate the arbitrary function $f \in L_T^2$ by the sequence of bounded function \( \{f_N\} \subset L_T^2 \) defined by

\[
    f_N(t, \omega) = \begin{cases} 
        N, & \text{if } f(t, \omega) > N, \\
        f(t, \omega), & \text{if } -N < f(t, \omega) < N, \\
        -N, & \text{if } f(t, \omega) < -N, 
    \end{cases} 
\]  \hspace{1cm} (3.20)

for some $N > 0$. The sequence $\{f_N\}_{N=1}^\infty$ is bounded by $N$. Moreover $|f_N(t, \omega)| \leq |f(t, \omega)|$ and $E[(f(t, \cdot))^2] < \infty$ for almost all $t \in [0, T]$, so by the theorem of dominated convergence

\[
    E\left[|f_N(t, \cdot) - f(t, \cdot)|^2\right] \to 0, 
\]  \hspace{1cm} (3.21)

for almost all $t \in [0, T]$. Furthermore

\[
    E\left[(f_N(t, \cdot) - f(t, \cdot))^2\right] \leq E\left[|f_N(t, \cdot)| + |f(t, \cdot)|\right]^2 \leq E\left[(2|f(t, \cdot)|)^2\right] \leq 4E\left[|f(t, \cdot)|^2\right], 
\]  \hspace{1cm} (3.22)

where $\int_0^T E\left[(f(t, \omega))^2\right] dt < \infty$, so another application of the theorem of dominated convergence asserts that

\[
    \int_0^T E\left[|f_N(t, \cdot) - f(t, \cdot)|^2\right] dt \to 0, \hspace{1cm} (3.23)
\]

as $N \to \infty$, that is $f_N$ converges to $f$ in $L_T^2$ (see the norm in (3.13)).

For each bounded function $f_N \in L_T^2$, we define the function $g_k$ by

\[
    g_k(t, \omega) = ke^{-kt} \int_0^t e^{ks} f_N(s, \omega) \, ds, \hspace{1cm} (3.24)
\]

for each $k > 0$. From the properties of $f_N$ and the fact that $g_k$ does not include values of $f_N$ at times exceeding $t$, it is clear that $g_k$ is jointly $\mathcal{L} \times \mathcal{A}$-measurable and $\mathcal{A}_t$-adapted. The bound on $f_N$ further implies that

\[
    |g_k(t, \omega)| \leq N \left(1 - e^{-kt}\right), \hspace{1cm} (3.25)
\]

so $E\left[(g_k(t, \cdot))^2\right]$ is finite for all $t \in [0, T]$ and integrable over $[0, T]$, implying that $g_k \in L_T^2$.

From the definition of $g_k$ and the properties of $f_N$ it can be shown that

\[
    |g_k(t, \omega) - g_k(s, \omega)| \leq 2kN(t - s), \hspace{1cm} (3.26)
\]

implying that the sample paths of $g_k$ are bounded and continuous. By construction, it is clear that $g_k \to f_N$ as $k \to \infty$, so two applications of the theorem of dominated convergence asserts that

\[
    \int_0^T E\left[|g_k(t, \cdot) - f_N(t, \cdot)|^2\right] dt \to 0, \hspace{1cm} (3.27)
\]

that is $g_k$ converges to $f_N$ in $L_T^2$. 

22
From (3.26) it also follows that
\[ E \left[ |g_k(t, \cdot) - g_k(t + h, \cdot)|^2 \right] \to 0, \tag{3.28} \]
as \( h \to 0 \) so \( g_k \) is mean-square continuous. Now consider partitions \( 0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_n^{(n)} = T \) such that \( \max_{0 \leq i \leq n-1} \left( t_{i+1}^{(n)} - t_i^{(n)} \right) \to 0 \) as \( n \to \infty \). Define sequences of step functions \( \{ h(n) \}_{n=1}^\infty \) as \( h(n)(t, \omega) = g_k(t_j, \omega) \) w.p.1 on \( t_j^{(n)} \leq t < t_{j+1}^{(n)} \) for \( 0 \leq j \leq n-1 \). Then it is clear that \( h(n) \in \mathcal{S}^2_T \). Furthermore
\[ E \left[ \left| h(n)(t, \cdot) - g_k(t, \cdot) \right|^2 \right] \leq 2E \left[ \left| h(n)(t, \cdot) - g_k(t_j, \cdot) \right|^2 \right] + 2E \left[ |g_k(t_j, \cdot) - g_k(t, \cdot)|^2 \right] \to 0, \tag{3.29} \]
as \( n \to \infty \). The first term vanishes because \( h(n)(t, \omega) = g_k(t_j, \omega) \) w.p.1. and \( E \left[ (g_k(t, \cdot))^2 \right] < \infty \), whereas the second term vanishes due to the mean-square continuity of \( g_k \). The theorem of dominated convergence together with the fact that \( E \left[ (g_k(t, \cdot))^2 \right] < \infty \) for all \( t \in [0, T] \) finally yields
\[ \int_0^T E \left[ \left| h(n)(t, \cdot) - g_k(t, \cdot) \right|^2 \right] dt \to 0, \tag{3.30} \]
showing that \( h(n) \) converges to \( g_k \) in \( \mathcal{L}^2_T \).

Since \( h(n) \to g_k \to f_N \to f \) in \( \mathcal{L}^2_T \), where \( h(n) \in \mathcal{S}^2_T \) it is clear that \( \mathcal{S}^2_T \) is dense in \( \mathcal{L}^2_T \). \( \blacksquare \)

For any function \( f \in \mathcal{L}^2_T \) Lemma 3.5 provides us with a sequence of step functions \( h(n) \in \mathcal{S}^2_T \) such that \( h(n) \) converges to \( f \) in \( \mathcal{L}^2_T \), that is
\[ \int_0^T E \left[ \left| h(n)(t, \cdot) - f(t, \cdot) \right|^2 \right] dt \to 0, \tag{3.31} \]
as \( n \to \infty \). By the linearity of the Itô integral \( I \left[ h(n) \right] \) for functions \( h(n) \in \mathcal{S}^2_T \) and the inequality \((a + b)^2 \leq 2a^2 + 2b^2\) we get
\[ E \left[ I \left[ h(n) \right] - I \left[ h(n+m) \right] \right]^2 = E \left[ I \left[ h(n) - h(n+m) \right] \right]^2 \]
\[ = \int_0^T E \left[ \left| h(n)(t, \cdot) - h(n+m)(t, \cdot) \right|^2 \right] dt \]
\[ \leq \int_0^T E \left[ \left| h(n)(t, \cdot) - f(t, \cdot) \right|^2 \right] dt + 2\int_0^T E \left[ \left| h(n+m)(t, \cdot) - f(t, \cdot) \right|^2 \right] dt, \tag{3.32} \]
which obviously turns to zero as \( m, n \to \infty \). Thus \( \{ I \left[ h(n) \right] \}_{n=1}^\infty \) is a Cauchy sequence in the complete space \( \mathcal{L}^2 (\Omega, \mathcal{F}, P) \) and there must exist a unique w.p.1 random variable \( I \in \mathcal{L}^2 (\Omega, \mathcal{F}, P) \) such that \( E \left[ I \left[ h(n) \right] - I \right]^2 \to 0 \) as \( n \to \infty \).

**Definition 3.6** The Itô integral of \( f \in \mathcal{L}^2_T \) is defined by
\[ \int_0^T f(t, \omega) \, dW_t(\omega) = \lim_{n \to \infty} \int_0^T h(n)(t, \omega) \, dW_t(\omega), \tag{3.33} \]
where the limit is in $L^2(\Omega, \mathcal{A}, P)$ and $\{h^{(n)}\}_{n=1}^{\infty}$ is a sequence of functions in $\mathcal{S}^2_T$ such that

$$
\int_0^T E \left[ |h^{(n)}(t, \cdot) - f(t, \cdot)|^2 \right] dt \to 0,
$$

as $n \to \infty$.

By the limiting procedure all properties of Itô integrals in $\mathcal{S}^2_T$ stated in Lemma 3.4 holds for Itô integrals in $\mathcal{L}^2_T$ as well. We demonstrate with an example how Itô integrals can be calculated.

**Example 3.7** Here we show that for the Wiener process $W$, we have

$$
\int_0^T W_s(\omega) dW_s(\omega) = \frac{1}{2} (W_T(\omega))^2 - \frac{1}{2} T,
$$

w.p.1. Note for comparison that if $w$ were a differentiable function satisfying $w(0) = 0$, the corresponding Riemann integral would be

$$
\int_0^T w(s) dw(s) = \int_0^T \left\{ \frac{1}{2} (w(s))^2 \right\} dt = \frac{1}{2} (w(T))^2,
$$

so the extra term $- \frac{T}{2}$ is a peculiarity of the Itô integral.

Put

$$
h^{(n)}(s, \omega) = \sum_{j=0}^{n-1} W_{i_j^{(n)}}(\omega) \chi_{[i_j^{(n)}, i_{j+1}^{(n)}]}(s).
$$

Obviously $h^{(n)} \in \mathcal{S}^2_T$. Introducing $\Delta t^{(n)} = \sup_{1 \leq j \leq n} (i_{j+1}^{(n)} - i_j^{(n)})$, we obtain

$$
\int_0^T E \left[ |h^{(n)}(t, \cdot) - W_t|^2 \right] dt
= \sum_{j=0}^{n-1} \int_{i_j^{(n)}}^{i_{j+1}^{(n)}} E \left[ |W_s - W_t|^2 \right] dt
= \sum_{j=0}^{n-1} \int_{i_j^{(n)}}^{i_{j+1}^{(n)}} E \left[ (\int_t^s dW_r)^2 \right] dt
= \sum_{j=0}^{n-1} \int_{i_j^{(n)}}^{i_{j+1}^{(n)}} \frac{(s-t)^2}{2} dt
\leq \frac{\Delta t^{(n)}}{2} \sum_{j=0}^{n-1} (i_{j+1}^{(n)} - i_j^{(n)})^2 = \frac{\Delta t^{(n)} T}{2} \to 0,
$$

as $\Delta t^{(n)} \to 0$. Here we used the Itô isometry in the third equality. Hence, since $h^{(n)} \to W$ in $\mathcal{L}^2_T$, we can use $h^{(n)}$ to define the Itô integral of $W$ and

$$
\int_0^T W_s(\omega) dW_s(\omega) = \lim_{n \to \infty} \int_0^T h^{(n)}(s, \omega) dW_s(\omega)
= \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} W_{i_j^{(n)}}(\omega) \left( W_{i_{j+1}^{(n)}}(\omega) - W_{i_j^{(n)}}(\omega) \right).
$$
Since

\[
(W_{t_{j+1}^{(n)}} (\omega))^2 - (W_{t_j^{(n)}} (\omega))^2 = (W_{t_{j+1}^{(n)}} (\omega) - W_{t_j^{(n)}} (\omega))^2 + 2W_{t_j^{(n)}} (\omega)(W_{t_{j+1}^{(n)}} (\omega) - W_{t_j^{(n)}} (\omega)),
\]

we get

\[
\sum_{j=0}^{n-1} W_{t_j^{(n)}} (\omega)(W_{t_{j+1}^{(n)}} (\omega) - W_{t_j^{(n)}} (\omega)) = \frac{1}{2} \sum_{j=0}^{n-1} \left( (W_{t_{j+1}^{(n)}} (\omega))^2 - (W_{t_j^{(n)}} (\omega))^2 \right) - \frac{1}{2} \sum_{j=0}^{n-1} (W_{t_{j+1}^{(n)}} (\omega) - W_{t_j^{(n)}} (\omega))^2 S_n
\]

\[
= \frac{1}{2} (W_T (\omega))^2 - \frac{1}{2} S_n, \tag{3.41}
\]

w.p.1. because the first sum is telescoping and \( W_0 = 0 \) w.p.1. It remains to show that \( S_n \to T \) in \( L^2 (\Omega, \mathcal{A}, P) \) as \( \Delta t^{(n)} \to 0 \). By the properties of the Wiener process we can easily deduce

\[
\lim_{\Delta t^{(n)} \to 0} E [S_n] = \lim_{\Delta t^{(n)} \to 0} E \left[ \sum_{j=0}^{n-1} (W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}})^2 \right] = \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} (t_{j+1}^{(n)} - t_j^{(n)}) = T, \tag{3.42}
\]

and, since \( W_{t_{j+1}^{(n)}} (\omega) - W_{t_j^{(n)}} (\omega) \) and \( W_{t_{j+1}^{(n)}} (\omega) - W_{t_k^{(n)}} (\omega) \) are independent for \( j \neq k \),

\[
\lim_{\Delta t^{(n)} \to 0} \text{Var} [S_n] = \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} \text{Var} \left[ (W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}})^2 \right]
\]

\[
= \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} \left[ E \left[ (W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}})^4 \right] - (t_{j+1}^{(n)} - t_j^{(n)})^2 \right]
\]

\[
= \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} \left[ 3 (t_{j+1}^{(n)} - t_j^{(n)})^2 - (t_{j+1}^{(n)} - t_j^{(n)})^2 \right]
\]

\[
\leq \lim_{\Delta t^{(n)} \to 0} 2\Delta t^{(n)} \sum_{j=0}^{n-1} (t_{j+1}^{(n)} - t_j^{(n)}) = \lim_{\Delta t^{(n)} \to 0} 2\Delta t^{(n)} T = 0, \tag{3.43}
\]

Here we have used the property \( E \left[ (W_t - W_s)^4 \right] = 3(t-s)^2 \) of the Wiener process. Since the variance approaches zero, the value of \( S_n \) converges in \( L^2 (\Omega, \mathcal{A}, P) \) to its mean value \( T \) as \( \Delta t^{(n)} \to 0 \).

This elementary result suggests that ordinary calculus rules cannot be applied to the Itô integral. In Section 3.4 we prove the Itô formula, which gives a set of calculus rules for Itô integrals and a means to calculate Itô integrals without using sequences of step functions.

Note that the Itô integral can be defined for more general classes of functions than \( \mathcal{L}_T^2 \) (compare for example the definition of \( \mathcal{L}_{T,T}^2 \) in [KP92]) but in these more general cases the resulting integrals will not always have as nice properties as in the case presented here.
3.3 Martingales

So far we have only considered the Itô integral on a fixed time interval $[0, T]$. We can easily generalize the domain of integration to Borel subsets $B$ of $[0, T]$ by multiplying the integrand $f$ by $\chi_B$, since clearly $f\chi_B \in \mathcal{L}^2$. For a variable subinterval $[t_0, t] \subset [0, T]$, it gets more cumbersome and, as a first step, we define the stochastic process $Z = \{Z_t : t_0 \leq t \leq T\}$ as

$$Z_t(\omega) = \int_{t_0}^{t} f(s, \omega) \, dW_s(\omega),$$  \hspace{1cm} (3.44)

w.p.1. In many ways $Z$ has properties that are similar to those of the Wiener process. In this section we shall show that this stochastic integral (as is the case for the Wiener process) is a martingale and has a continuous version.

**Definition 3.8** Let $M$ be a stochastic process and $\{\mathcal{M}_t\}$ a filtration (increasing family of $\sigma$-algebras) to which $M$ is adapted. Then $M$ is said to be a martingale with respect to the filtration $\{\mathcal{M}_t\}$ if $E[|M_t|] < \infty$ and

$$E[M_t|M_s] = M_s,$$  \hspace{1cm} (3.45)

for all $t \geq s$.

We showed in (3.11) that the Wiener process is a martingale with respect to the filtration $\{A_t\}$ generated by the Wiener process. In fact it is possible to define stochastic integrals with the Wiener process replaced by any martingale.

**Proposition 3.9 (3.2.5 in [KP92])** If $f \in \mathcal{L}^2$, the stochastic integral $Z$ defined in (3.44) is a martingale with respect to the filtration $\{A_t\}$.

**Proof.** Since $f \in \mathcal{L}^2$ it follows that

$$(E[Z_t])^2 \leq E[Z_t^2] = \int_{t_0}^{t} E\left([f(s, \cdot)]^2\right) \, ds < \int_{0}^{T} E\left([f(s, \cdot)]^2\right) \, ds < \infty.$$  \hspace{1cm} (3.46)

For property (3.45), note that it follows from the construction of the Itô integral that for general $f \in \mathcal{L}^2$, there exists a sequence $f^{(n)} \in \mathcal{S}^2$ such that $Z^{(n)}$ corresponding to $f^{(n)}$ converges to $Z$ in $\mathcal{L}^2(\Omega, \mathcal{A}, \mathbb{P})$. It thus suffices to show (3.45) for $f^{(n)} \in \mathcal{S}^2$. Let $0 \leq s \leq t \leq T$ and introduce a partition $s = t^{(n)}_0 \leq t^{(n)}_1 \leq \ldots \leq t^{(n)}_n = t$ of the interval $[s, t]$. Furthermore let $f^{(n)}(t, \omega) = f^{(n)}_j(\omega)$ on $t^{(n)}_j \leq t < t^{(n)}_{j+1}$ so that $f^{(n)}(t, \omega)$ is $A_t$-measurable. Then

$$E\left[Z^{(n)}_t - Z^{(n)}_s\right] = \sum_{j=0}^{n-1} E\left[f^{(n)}_j\left(W^{(n)}_{t^{(n)}_{j+1}} - W^{(n)}_{t^{(n)}_j}\right)\right] | A_s$$

$$= \sum_{j=0}^{n-1} E\left[f^{(n)}_j\left(W^{(n)}_{t^{(n)}_{j+1}} - W^{(n)}_{t^{(n)}_j}\right)\right] | A_s$$

$$= \sum_{j=0}^{n-1} \left[f^{(n)}_jE\left(W^{(n)}_{t^{(n)}_{j+1}} - W^{(n)}_{t^{(n)}_j}\right)\right] | A_s = 0,$$  \hspace{1cm} (3.47)

whereby (3.45) easily follows. $\blacksquare$
Since stochastic integrals are martingales they satisfy the martingale inequality and the Doob inequality below.

**Proposition 3.10** For a positive martingale $Y$ it holds that

$$\mathbb{P} \left( \sup_{0 \leq t \leq T} Y_t > r \right) \leq \frac{1}{r^p} \mathbb{E} \left[ |Y_T^p| \right].$$

(3.48)

for any $r > 0$ and $p \geq 1$. If $f \in \mathcal{L}_T^2$ it immediately follows that the stochastic integral $Z$ defined in (3.44) satisfies

$$\mathbb{P} \left( \sup_{t_0 \leq s \leq t} Z_s > r \right) \leq \frac{1}{r^2} \mathbb{E} \left[ |Z_t - Z_{t_0}|^2 \right] = \frac{1}{r^2} \int_{t_0}^t \mathbb{E} \left[ (f(s, \cdot))^2 \right] \, ds.$$

(3.49)

**Proposition 3.11** For a positive martingale $Y$ it holds that

$$\mathbb{E} \left( \sup_{0 \leq t \leq T} |Y_t|^r \right) \leq \left( \frac{r}{r - 1} \right) \mathbb{E} \left[ |Y_T|^r \right],$$

(3.50)

for any $r > 1$. If $f \in \mathcal{L}_T^2$ it immediately follows that the stochastic integral $Z$ defined in (3.44) satisfies

$$\mathbb{E} \left( \sup_{t_0 \leq s \leq t} |Z_s|^2 \right) \leq \left( \frac{2}{2 - 1} \right)^2 \mathbb{E} \left[ |Z_t - Z_{t_0}|^2 \right] = 4 \int_{t_0}^t \mathbb{E} \left[ (f(s, \cdot))^2 \right] \, ds.$$

(3.51)

We now turn to the second important result of this section.

**Theorem 3.12** (3.2.6 in [KP92]) Let $Z$ be the stochastic integral defined by (3.44) and suppose that $f \in \mathcal{L}_T^2$. Then there exists a continuous version of $Z$.

**Proof.** For $f \in \mathcal{L}_T^2$, let $f^{(n)} \in \mathcal{S}_T^2$ be a sequence such that $Z^{(n)}$ corresponding to $f^{(n)}$ converges to $Z$ in $L^2(\Omega, \mathcal{A}, \mathbb{P})$. Then, since both $Z^{(n)}$ and $Z$ are martingales, so is $Z - Z^{(n)}$ and we can apply the martingale inequality (Proposition 3.10) to this difference. Hence

$$\mathbb{P} \left( \sup_{t_0 \leq s \leq T} |Z_s - Z^{(n)}_s| > \frac{1}{n} \right) \leq n^2 \int_{t_0}^T \mathbb{E} \left[ \left| f(s, \cdot) - f^{(n)}(s, \cdot) \right|^2 \right] \, ds \leq \frac{1}{n^2},$$

(3.52)

if we choose the step functions so that

$$\int_{t_0}^T \mathbb{E} \left[ \left| f(s, \cdot) - f^{(n)}(s, \cdot) \right|^2 \right] \, ds \leq \frac{1}{n^4},$$

(3.53)

for $n = 1, 2, \ldots$ Hence, the infinite series

$$\sum_{n=1}^\infty \mathbb{P} \left( \sup_{t_0 \leq s \leq T} |Z_s - Z^{(n)}_s| > \frac{1}{n} \right) \leq \sum_{n=1}^\infty \frac{1}{n^2} < \infty,$$

(3.54)

is convergent and by the Borel-Cantelli Lemma

$$\mathbb{P} \left( \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty \sup_{t_0 \leq s \leq T} |Z_s - Z^{(k)}_s| > \frac{1}{k} \right) = 0,$$

(3.55)

27
so the probability that \( \sup_{t_0 \leq s \leq T} |Z_s - Z_s^{(k)}| > \frac{1}{k} \) for infinitely many \( k \) is zero. We conclude that for almost all \( \omega \in \Omega \) there must exist a \( k_1(\omega) \) such that

\[
\sup_{t_0 \leq s \leq T} |Z_s - Z_s^{(k)}| \leq \frac{1}{k}, \quad \text{for all } k > k_1(\omega),
\]

(3.56)
or, in other words, that

\[
\lim_{k \to \infty} \sup_{t_0 \leq s \leq T} |Z_s - Z_s^{(k)}| = 0
\]

(3.57)

for almost all \( \omega \in \Omega \), showing that the sequence of stochastic integrals \( Z^{(n)} \) is uniformly convergent on \([t_0, T]\) w.p.1. Each of the stochastic integrals \( Z^{(n)} \) is continuous. Indeed, this follows because if we choose \( s \) and \( t \) so close that they fall within the same partition interval the difference \( Z^{(n)}_t - Z^{(n)}_s \) equals a constant \( f^{(n)}_i \) times \( W_t - W_s \), which, using the continuity of the Wiener process, approaches zero when \( t \to s \). Since the uniform limit of a sequence of continuous functions is continuous we conclude that \( Z \) is continuous w.p.1.

### 3.4 Itō’s formula

Up to this point, calculating stochastic integrals has been quite an elaborate task. This will change now as we derive Itō’s formula, which simplifies the stochastic calculus considerably.

**Definition 3.13** Let \((\Omega, \mathcal{A}, P)\) be a probability space. Then a stochastic differential (or Itō process) is an expression on the form

\[
X_t(\omega) = X_s(\omega) + \int_s^t e_u(\omega) \, du + \int_s^t f_u(\omega) \, dW_u(\omega),
\]

(3.58)

w.p.1 for any \( 0 \leq s \leq t \leq T \), where the functions \( e \) and \( f \) are jointly \( \mathcal{L} \times \mathcal{A} \)-measurable, \( \mathcal{A}_t \)-adapted and satisfies, w.p.1., the criteria

\[
\int_s^t |e_u(\omega)| \, du < \infty,
\]

(3.59)

\[
\int_s^t (f_u(\omega))^2 \, du < \infty.
\]

(3.60)

A stochastic differential is a stochastic process that can formally be written as

\[
dX_t = e_t \, dt + f_t \, dW_t.
\]

(3.61)

The sample paths of stochastic differentials are continuous w.p.1 (this is a consequence of Theorem 3.12). We shall now state and prove the Itō formula, which asserts that the space of stochastic differentials is closed under mappings in \( C^{1,2} \) and gives us a means to calculate functions of stochastic differentials.

**Theorem 3.14 (Itō’s formula, 3.3.2 in [KP92])** Let \( X_t \) be a stochastic differential and let

\[
U : [0, T] \times \mathbb{R} \to \mathbb{R}
\]

be a function with continuous partial derivatives \( \frac{\partial U}{\partial t}, \frac{\partial U}{\partial x} \) and \( \frac{\partial^2 U}{\partial x^2} \).

Then \( Y_t = U(t, X_t) \) is again a stochastic differential and

\[
Y_t - Y_s = \int_s^t \left\{ \frac{\partial U}{\partial t}(u, X_u) + e_u \frac{\partial U}{\partial x}(u, X_u) + \frac{1}{2} f_u^2 \frac{\partial^2 U}{\partial x^2}(u, X_u) \right\} \, du + \int_s^t f_u \frac{\partial U}{\partial x}(u, X_u) \, dW_u,
\]

(3.62)
w.p.1 for any 0 ≤ s ≤ t ≤ T.

**Proof.** The functions e and f can be approximated by sequences of step functions \( \{e^{(n)}\} \) and \( \{f^{(n)}\} \) in \( S^2_T \) such that

\[
\int_s^t |e_u^{(n)}(\omega) - e_u(\omega)| \, du < \infty, \quad \int_s^t \left( f_u^{(n)}(\omega) - f_u(\omega) \right)^2 \, du < \infty, \tag{3.63}
\]

If the result is shown for \( e^{(n)} \) and \( f^{(n)} \) one can show mean-square convergence to zero as

\[
\text{w.p.1}
\]

for \( n \rightarrow \infty \). Hence we prove the theorem in the special case when \( e \) and \( f \) are step functions and initially we suppose that \( e \) and \( f \) are constant functions. Let \( s = t_0^{(n)} < t_1^{(n)} < \ldots < t_n^{(n)} = t \) be a partition with \( \Delta t_j^{(n)} = t_{j+1}^{(n)} - t_j^{(n)} \to 0 \). We can expand the difference \( Y_t - Y_s \) as

\[
Y_t - Y_s = U(t, X_t) - U(s, X_s) = \sum_{j=0}^{n-1} \Delta U_j^{(n)}, \tag{3.64}
\]

where

\[
\Delta U_j^{(n)} = U\left(t_{j+1}^{(n)}, X_{t_{j+1}^{(n)}}\right) - U\left(t_j^{(n)}, X_{t_j^{(n)}}\right), \tag{3.65}
\]

for \( j = 0, \ldots, n-1 \). By the Taylor and Mean Value theorems there exist constants \( 0 \leq \alpha_j^{(n)} \leq 1 \) and \( 0 \leq \beta_j^{(n)} \leq 1 \) such that

\[
\Delta U_j^{(n)} = \frac{\partial U}{\partial t}\left(t_j^{(n)} + \alpha_j^{(n)} \Delta t_j^{(n)}, X_{t_j^{(n)}}\right) \Delta t_j^{(n)} + \frac{\partial U}{\partial x}\left(t_j^{(n)}, X_{t_j^{(n)}}\right) \Delta X_{t_j^{(n)}}
+ \frac{1}{2} \frac{\partial^2 U}{\partial x^2}\left(t_j^{(n)}, X_{t_j^{(n)}} + \beta_j^{(n)} \Delta X_{t_j^{(n)}}\right) \left(\Delta X_{t_j^{(n)}}\right)^2, \tag{3.66}
\]

w.p.1, where \( \Delta X_{t_j^{(n)}} = X_{t_{j+1}^{(n)}}^{(n)} - X_{t_j^{(n)}}^{(n)} \) for \( j = 0, \ldots, n-1 \). By the continuity of \( \frac{\partial U}{\partial t} \) and \( \frac{\partial^2 U}{\partial x^2} \) and the w.p.1 continuity of \( X_t \) we obtain

\[
\frac{\partial U}{\partial t}\left(t_j^{(n)} + \alpha_j^{(n)} \Delta t_j^{(n)}, X_{t_j^{(n)}}\right) \to \frac{\partial U}{\partial t}\left(t_j^{(n)}, X_{t_j^{(n)}}\right), \text{ w.p.1}, \tag{3.67}
\]

and

\[
\frac{\partial^2 U}{\partial x^2}\left(t_j^{(n)}, X_{t_j^{(n)}} + \beta_j^{(n)} \Delta X_{t_j^{(n)}}\right) \to \frac{\partial^2 U}{\partial x^2}\left(t_j^{(n)}, X_{t_j^{(n)}}\right), \text{ w.p.1}, \tag{3.68}
\]

as \( \Delta t^{(n)} = \max_{0 \leq j \leq n-1} \Delta t_j^{(n)} \to 0 \). Furthermore, for constant \( e \) and \( f \), the increment \( \Delta X_{t_j^{(n)}} \) can be written \( \Delta X_{t_j^{(n)}} = e \Delta t_j^{(n)} + f \Delta W_{t_j^{(n)}}^{(n)} \), where \( \Delta W_{t_j^{(n)}}^{(n)} = W_{t_{j+1}^{(n)}}^{(n)} - W_{t_j^{(n)}}^{(n)} \) for \( j = 0, \ldots, n-1 \), so

\[
\sum_{j=0}^{n-1} \left\{ \left(\Delta X_{t_j^{(n)}}\right)^2 - \left(f \Delta W_{t_j^{(n)}}^{(n)}\right)^2 \right\} = e^2 \sum_{j=0}^{n-1} \left(\Delta t_j^{(n)}\right)^2 + 2ef \sum_{j=0}^{n-1} \Delta W_{t_j^{(n)}}^{(n)} \Delta t_j^{(n)} = e^2 \Delta t^{(n)} + 2ef \sum_{j=0}^{n-1} \Delta W_{t_j^{(n)}}^{(n)} \Delta t_j^{(n)}. \tag{3.69}
\]

It is obvious that the first term on the right hand side of (3.69) tends to zero w.p.1 as \( \Delta t^{(n)} \to 0 \).

For the second term we can show mean-square convergence to zero

\[
E \left[ \left(2ef \sum_{j=0}^{n-1} \Delta W_{t_j^{(n)}}^{(n)} \Delta t_j^{(n)}\right)^2 \right] = 4(ef)^2 \sum_{j=0}^{n-1} \left(\Delta t_j^{(n)}\right)^3 \to 0, \tag{3.70}
\]
as $\Delta t^{(n)} \to 0$, which implies convergence in probability.

Combining (3.64) and (3.66)-(3.69) we obtain

\[
Y_t - Y_s = \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} \Delta t_j^{(n)} = \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} \left\{ \frac{\partial U}{\partial t} + e \frac{\partial U}{\partial x} + \frac{1}{2} f^2 \frac{\partial^2 U}{\partial x^2} \right\} \left( t_j^{(n)}, X_{t_j^{(n)}} \right) \Delta t_j^{(n)} \\
+ \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} \left\{ f \frac{\partial U}{\partial x} \right\} \left( t_j^{(n)}, X_{t_j^{(n)}} \right) \Delta W_{t_j^{(n)}},
\]

where the limits are taken in probability. The first two terms in (3.61) are exactly the terms on the right hand side of (3.62), so it remains to show that the last limit in (3.71) is zero.

Let $\gamma_j^{(n)} = \left( \Delta W_{t_j^{(n)}} \right)^2 - \Delta t_j^{(n)}$ and let $I_{j,N}^{(n)}$ denote the indicator function of the events

\[ A_{j,N}^{(n)} = \{ \omega \in \Omega : X_{t_i^{(n)}} \leq N \text{ for } i = 0, \ldots, j \}, \tag{3.72} \]

for $j = 0, \ldots, n - 1$. For fixed $n$ the random variables $\gamma_j^{(n)}$ are independent and satisfy

\[ E \left[ \gamma_j^{(n)} \right] = 0, \tag{3.73} \]

\[ E \left[ \gamma_j^{(n)} \right]^2 = E \left( \Delta W_{t_j^{(n)}} \right)^4 - 2 \left( \Delta W_{t_j^{(n)}} \right)^2 \Delta t_j^{(n)} + \left( \Delta t_j^{(n)} \right)^2 \]

\[ = 3 \left( \Delta t_j^{(n)} \right)^2 - 2 \left( \Delta t_j^{(n)} \right)^2 + \left( \Delta t_j^{(n)} \right)^2 = 2 \left( \Delta t_j^{(n)} \right)^2. \tag{3.74} \]

From this we conclude

\[ E \left[ \sum_{j=0}^{n-1} \frac{\partial^2 U}{\partial x^2} \left( t_j^{(n)}, X_{t_j^{(n)}} \right) I_{j,N}^{(n)} \gamma_j^{(n)} \right]^2 = \sum_{j=0}^{n-1} E \left[ \left( \frac{\partial^2 U}{\partial x^2} \left( t_j^{(n)}, X_{t_j^{(n)}} \right) I_{j,N}^{(n)} \gamma_j^{(n)} \right)^2 \right] \leq C \sum_{j=0}^{n-1} 2 \left( \Delta t_j^{(n)} \right)^2 \leq 2C (t - s) \Delta t^{(n)} \to 0, \tag{3.75} \]

where $C = \max_{0 \leq s \leq t} \left| \frac{\partial^2 U}{\partial x^2} (s, x) \right|^2$. Since mean-square convergence implies convergence in probability, this establishes that last limit in (3.71) converges to zero in probability provided that

\[ P \left( I_{j,N}^{(n)} = 1 \right) \to 1, \tag{3.76} \]

as $N \to \infty$. But it is clear that

\[ \bigcup_{j=0}^{n-1} \left( A_{j,N}^{(n)} \right)^c \subseteq \left\{ \omega \in \Omega : \sup_{t \leq u \leq t} |X_u| > N \right\}, \tag{3.77} \]
Theorem 2.9c in [Gar88] states that for arbitrary numbers \( r \) and \( N \) the stochastic integral \( Z_t = \int_s^t f(u) \, dW_u \) satisfies the inequality

\[
P \left( \sup_{s \leq u \leq t} |Z_t| > N \right) \leq P \left( \int_{t_0}^t (f(s))^2 \, ds > r \right) + \frac{r}{N^2}.
\]

(3.78)

Generalizing this inequality to \( X_t \) it follows that

\[
P \left( \bigcup_{j=0}^{n-1} \left( A_{j,N}^{(n)} \right)^c \right) \to 0
\]

(3.79)

as \( N \to \infty \) because \( e \) and \( f \) are constants. This proves (3.76) and we have shown that (3.62) is valid with convergence in probability.

Now if a sequence of random variables converges in probability, there exists a subsequence w.p.1 (see Section 1.4 in [Gar88]) so by considering this subsequence, it is clear that (3.62) holds w.p.1. The proof is similar when \( e \) and \( f \) are step functions because if we make the partition sufficiently fine, the values of \( e \) and \( f \) are constant within the partition subintervals.

The Itô formula gives a first-order expansion for functions of stochastic differentials. Comparing with a first-order Taylor expansion, the term \( \frac{1}{2} \int_a^t \frac{\partial^2 U}{\partial x^2} (u, X_u) \) is unexpected. This term emerges since the differential \((dW_t)^2\) behaves like \( dt \) in the Itô calculus (as we saw in the proof of Itô’s formula) and it must thus be taken into account in a first-order expansion.

Formally the Itô formula can be written in differential form as

\[
dY_t = \frac{\partial U}{\partial t} (t, X_t) \, dt + \frac{\partial U}{\partial x} (t, X_t) \, dX_t + \frac{1}{2} \frac{\partial^2 U}{\partial x^2} (t, X_t) (dX_t)^2,
\]

(3.80)

where \((dX_t)^2\) can be calculated according to the rules

\[
(dt)^2 = dt \, dW_t = dW_t \, dt = 0, \quad (dW_t)^2 = dt.
\]

(3.81)

We use this version of the formula in the following example.

Example 3.15 The integral

\[
\int_0^T W_t \, dW_t = \frac{1}{2} W_T^2 - \frac{1}{2} T,
\]

(3.82)

was calculated in Example 3.7. Here we derive it by means of Itô’s formula.

Let \( U(t, x) = \frac{1}{2} x^2 \) and consider the stochastic differential \( X_t \) defined by

\[
dX_t = dW_t,
\]

(3.83)

so that \( X_t \) equals \( W_t \). Then \( Y_t \) is given by

\[
Y_t = U(t, X_t) = U(t, W_t) = \frac{1}{2} W_t^2,
\]

(3.84)

and by Itô’s formula we get

\[
dY_t = \frac{\partial U}{\partial t} (t, W_t) \, dt + \frac{\partial U}{\partial x} (t, W_t) \, dW_t + \frac{1}{2} \frac{\partial^2 U}{\partial x^2} (t, W_t) (dW_t)^2 = 0 + W_t \, dW_t + \frac{1}{2} (dW_t)^2
\]

\[
= W_t \, dW_t + \frac{1}{2} \, dt,
\]

(3.85)
where we have used $\frac{\partial U}{\partial t} = 0$, $\frac{\partial U}{\partial x} = x$ and $\frac{\partial^2 U}{\partial x^2} = 1$. Hence

$$d \left( \frac{1}{2} W_t^2 \right) = W_t \, dW_t + \frac{1}{2} \, dt,$$

or, in other words,

$$\frac{1}{2} W_T^2 = \int_0^T d \left( \frac{1}{2} W_t^2 \right) = \int_0^T W_t \, dW_t + \int_0^T \frac{1}{2} \, dt = \int_0^T W_t \, dW_t + \frac{1}{2} T. \quad (3.87)$$

Rearranging the terms yields

$$\int_0^T W_t \, dW_t = \frac{1}{2} W_T^2 - \frac{1}{2} T, \quad (3.88)$$

as was previously derived.

3.5 Vector valued stochastic differentials

It is straightforward to generalize Itô’s formula to vector valued stochastic differentials. In analogy with the properties shown for the one-dimensional Wiener process right after Definition 3.1, we define a multi-dimensional Wiener process as follows.

**Definition 3.16** An $m$-dimensional Wiener process $W = (W^{(1)},...,W^{(m)})$ with independent components associated with the filtration $\mathcal{F}^{(m)}_t : t \geq 0$ consists of $m$ one-dimensional Wiener processes with the properties

1. $W^{(k)}_t$ is $\mathcal{F}^{(m)}_0$-measurable,
2. $E \left[ W^{(k)}_t \big| \mathcal{F}^{(m)}_0 \right] = 0,$
3. $E \left[ \left( W^{(k)}_t - W^{(k)}_s \right) \left( W^{(l)}_t - W^{(l)}_s \right) \big| \mathcal{F}^{(m)}_s \right] = (t-s) \delta_{kl},$

w.p.1 for $0 \leq s \leq t$ and $1 \leq k, l \leq m.$

Let $e : [0,T] \times \Omega \to \mathbb{R}^d$ and $F : [0,T] \times \Omega \to \mathbb{R}^{d \times m}$ be vector and matrix functions, respectively. Assume that the components of $e$ and $F$ satisfy the criteria in the definition of scalar stochastic differentials. Then we can define the $d$-dimensional stochastic differential $X_t$ as

$$X^{(i)}_t (\omega) - X^{(i)}_s (\omega) = \int_s^t e^{(i)}_u (\omega) \, du + \sum_{k=1}^m \int_s^t F^{(i,k)}_u (\omega) \, dW^{(k)}_u (\omega). \quad (3.89)$$

w.p.1 for $i = 1,...,d$ and any $0 \leq s \leq t \leq T$. We state Itô’s formula for such stochastic differentials.

**Theorem 3.17** Let $X_t$ be a $d$-dimensional stochastic differential. Let $U : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be a function with continuous partial derivatives $\frac{\partial U}{\partial t}, \frac{\partial U}{\partial x_k}$ and $\frac{\partial^2 U}{\partial x_k \partial x_l}$. Then $Y_t = U (t, X_t) = \quad \text{\textit{32}}$
U \left( t, X_t^{(1)}, \ldots, X_t^{(d)} \right) is again a stochastic differential and

\[ Y_t - Y_s = \int_s^t \left\{ \frac{\partial U}{\partial t} + \sum_{i=1}^d e_u^{(i)} \frac{\partial U}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d \sum_{k=1}^m F_u^{(i,k)} F_u^{(j,k)} \frac{\partial^2 U}{\partial x_i \partial x_j} \right\} du + \int_s^t \sum_{i=1}^d \sum_{k=1}^m F_u^{(i,k)} \frac{\partial U}{\partial x_i} dW^{(k)}_u, \tag{3.90} \]

w.p.1 for any \( 0 \leq s \leq t \leq T \), where all functions and partial derivatives are evaluated at \((u, X_u)\).

Formally this result can be written in differential form as

\[ dY_t = \frac{\partial U}{\partial t} dt + \sum_{i=1}^d \frac{\partial U}{\partial x_i} dX_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 U}{\partial x_i \partial x_j} dX_t^{(i)} dX_t^{(j)}, \tag{3.91} \]

where \( dX_t^{(i)} dX_t^{(j)} \) can be calculated according to the rules

\[ (dt)^2 = dt \, dW_t^{(k)} = dW_t^{(k)} dt = 0, \quad dW_t^{(k)} dW_t^{(l)} = \delta_{kl} dt. \tag{3.92} \]

The proof is analogous to the one-dimensional case.

### 3.6 The Stratonovich integral

As we discovered in Section 3.1, different choices of evaluation points give rise to different stochastic integrals. When using stochastic differentials for modelling of physical phenomena, it is not obvious which choice of evaluation point that leads to the most reasonable model. One benefit of choosing the Itô interpretation is that the integrands may not "look into the future", which seem suitable for many applications, for example those in biology and finance.

The most common alternative choice is the Stratonovich integral

\[ \int_0^T f(s, \omega) \circ dW_s(\omega), \tag{3.93} \]

which is a linear combination of the stochastic integrals evaluated at the left and right end points. Adjusting the results in Example 3.7 to the Stratonovich case, we deduce that

\[ \int_0^T W_s(\omega) \circ dW_s(\omega) = \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} \frac{W_{t_j}^{(n)} + W_{t_{j+1}}^{(n)}}{2} \left( W_{t_{j+1}}^{(n)} - W_{t_j}^{(n)} \right) \]

\[ = \lim_{\Delta t^{(n)} \to 0} \sum_{j=0}^{n-1} \frac{W_{t_j}^{(n+1)} - W_{t_j}^{(n)}}{2} = \frac{W_T^2}{2}. \tag{3.94} \]

As demonstrated in (3.36), this result agrees with ordinary calculus. It can be shown that the Stratonovich integral obeys all ordinary calculus rules (the extra term found in Itô’s formula vanishes in the Stratonovich case) and this is of course the main advantage of this integral. On the other hand, the Stratonovich integral is not a martingale and does not satisfy the neat expressions for the moments (see Proposition 3.4) that hold for the Itô integral.
4 Stochastic differential equations

Equipped with the definition of the Itô integral, the stochastic integral equation presented in (3.3),
\[ X_t(\omega) = X_0(\omega) + \int_0^t a(s, X_s(\omega)) \, ds + \int_0^t b(s, X_s(\omega)) \, dW_s(\omega), \tag{4.1} \]
now makes sense. We can write (4.1) in differential form as
\[ dX_t = a(t, X_t) \, dt + b(t, X_t) \, dW_t, \tag{4.2} \]
and call such an equation a stochastic differential equation (SDE). It is natural to ask the following two questions:

1. Can one obtain existence and uniqueness of solutions to stochastic differential equations (and for which prerequisites)? What are the properties of the solutions?
2. How can one find an explicit solution to a given stochastic differential equation?

We address these two questions in reverse order.

4.1 Analytical solutions

We begin with a simple but important example that was discussed in Section 1.1.

Example 4.1 Let \( N_t \) denote the size of a population at time \( t \). The stochastic counterpart to the exponential growth model for \( N_t \)
\[ \frac{dN_t}{dt} = r(t) N_t, \tag{4.3} \]
is
\[ dN_t = e(t) \, dt + f(t) \, dW_t, \tag{4.4} \]
previously stated in (1.4). The solution to this equation is known as the geometric Brownian motion. We shall derive an explicit expression for the solution \( N_t \).

Dividing both sides of (4.4) by \( N_t \) and integrating both sides yields
\[ \frac{dN_t}{N_t} = e(t) \, dt + f(t) \, dW_t \Rightarrow \int_0^T \frac{dN_t}{N_t} = \int_0^T e(t) \, dt + \int_0^T f(t) \, dW_t. \tag{4.5} \]
Guided by the form of the integral term on the left hand side of the second equality in (4.5), we apply Itô’s formula to the function
\[ U(t, x) = \log x, \tag{4.6} \]
and obtain
\[ d(\log N_t) = \frac{\partial U}{\partial t}(t, N_t) \, dt + \frac{\partial U}{\partial x}(t, N_t) \, dN_t + \frac{1}{2} \frac{\partial^2 U}{\partial x^2}(t, N_t) (dN_t)^2 \tag{4.7} \]
\[ = 0 + \frac{1}{N_t} \, dN_t + \frac{1}{2} \left( \frac{1}{N_t^2} \right) (dN_t)^2 = \frac{dN_t}{N_t} - \frac{(f(t) N_t)^2}{2N_t^2} \, dt = \frac{dN_t}{N_t} - \frac{(f(t))^2}{2} \, dt. \]
Hence
\[ \int_0^T \frac{dN_t}{N_t} = \int_0^T d(\log N_t) + \int_0^T \frac{(f(t))^2}{2} \, dt = \log \frac{N_T}{N_0} + \frac{1}{2} \int_0^T (f(t))^2 \, dt. \tag{4.8} \]
Combining (4.5) and (4.8), we get
\[
\log \frac{N_T}{N_0} = \int_0^T e(t) \, dt + \int_0^T f(t) \, dW_t - \frac{1}{2} \int_0^T (f(t))^2 \, dt, \tag{4.9}
\]
or
\[
N_T = N_0 \exp \left( \int_0^T \left( e(t) - \frac{1}{2} (f(t))^2 \right) \, dt + \int_0^T f(t) \, dW_t \right), \tag{4.10}
\]
which in the special case of constant \(e\) and \(f\) reduces to
\[
N_T = N_0 \exp \left( \left( e - \frac{1}{2} f^2 \right) T + fW_T \right). \tag{4.11}
\]

If \(W_t\), for \(t \leq T\), is independent of \(N_0\) we expect the expectation of \(N_T\) to coincide with the solution to the deterministic exponential growth model or, in other words, we expect \(E[N_T] = E[N_0] \exp \left( \int_0^T e(t) \, dt \right)\) to hold. We shall show that this is indeed the case. To investigate the expression \(E \left[ \exp \left( \int_0^T f(t) \, dW_t \right) \right]\), we let \(Y_t = \exp \left( X_t \right)\) with
\[
dX_t = f(t) \, dW_t + \frac{1}{2} \exp(X_t)(f(t))^2 \, dt,
\]
or
\[
Y_T - Y_0 = \frac{1}{2} \int_0^T \exp(X_t)(f(t))^2 \, dt + \int_0^T \exp(X_t) f(t) \, dW_t. \tag{4.13}
\]
Since the expectation of an Itô integral is zero, we obtain
\[
E[Y_T] = E[Y_0] + \frac{1}{2} \int_0^T (f(t))^2 \, E[Y_t] \, dt, \tag{4.14}
\]
or in differential form
\[
\frac{dE[Y_T]}{dT} = \frac{1}{2} (f(T))^2 E[Y_T], \quad E[Y_0] = 1, \tag{4.15}
\]
with the solution
\[
E \left[ \exp \left( \int_0^T f(t) \, dW_t \right) \right] = E[Y_T] = \exp \left( \frac{1}{2} \int_0^T (f(t))^2 \, dt \right). \tag{4.16}
\]
Finally (4.10) and (4.16) yield
\[
E[N_T] = E \left[ N_0 \exp \left( \int_0^T \left( e(t) - \frac{1}{2} (f(t))^2 \right) \, dt + \int_0^T f(t) \, dW_t \right) \right]
= E[N_0] \exp \left( \int_0^T f(t) \, dW_t \right) \exp \left( \int_0^T \left( e(t) - \frac{1}{2} (f(t))^2 \right) \, dt \right)
= E[N_0] \exp \left( \int_0^T e(t) \, dt \right), \tag{4.17}
\]
as desired.
The geometric Brownian motion is a special case of a scalar linear stochastic differential equation

\[ dX_t = (a_1(t) X_t + a_2(t)) \, dt + (b_1(t) X_t + b_2(t)) \, dW_t, \quad (4.18) \]

with \( a_2(t) \equiv b_2(t) \equiv 0, \quad a_1(t) = e(t) \) and \( b_1(t) = f(t) \). When the functions \( a_1, a_2, a_3 \) and \( a_4 \) are Lebesgue measurable and bounded on an interval \([0, T]\), Theorem 4.4 below asserts the existence of a unique solution to this equation.

We shall use the solution to the geometric Brownian motion in the derivation of a solution to the general linear stochastic differential equation. We introduce the notation

\[ \Phi_{0,t} = \exp \left( \int_0^t \left( a_1(u) - \frac{1}{2} (b_1(u))^2 \right) \, du + \int_0^t b_1(u) \, dW_u \right), \quad (4.19) \]

for the solution to geometric Brownian motion with initial value one. Applying Itô’s formula on the function

\[ \Phi_{0,t}^{-1} = \exp \left( \int_0^t \left( -a_1(u) + \frac{1}{2} (b_1(u))^2 \right) \, du - Y_t \right), \quad (4.20) \]

where \( dY_t = b_1(t) \, dW_t \), we obtain

\[
\begin{align*}
\frac{d}{dt} \left( \Phi_{0,t}^{-1} \right) &= \Phi_{0,t}^{-1} \left( -a_1(t) + \frac{1}{2} (b_1(t))^2 \right) \, dt - \Phi_{0,t}^{-1} b_1(t) \, dW_t + \frac{1}{2} \Phi_{0,t}^{-1} (b_1(t))^2 \, dt \\
&= \Phi_{0,t}^{-1} \left( -a_1(t) + (b_1(t))^2 \right) \, dt - \Phi_{0,t}^{-1} b_1(t) \, dW_t.
\end{align*}
\]

We can use the integration by parts formula (see Exercise 24) on the function \( \Phi_{0,t}^{-1} X_t \), where \( X_t \) satisfies (4.18). This yields

\[
\begin{align*}
\frac{d}{dt} \left( \Phi_{0,t}^{-1} X_t \right) &= \Phi_{0,t}^{-1} dX_t + X_t \frac{d}{dt} \left( \Phi_{0,t}^{-1} \right) + d \left( \Phi_{0,t}^{-1} \right) \, dX_t \\
&= \Phi_{0,t}^{-1} ((a_1(t) X_t + a_2(t)) \, dt + (b_1(t) X_t + b_2(t)) \, dW_t) \\
&\quad + X_t \left( \Phi_{0,t}^{-1} \left( -a_1(t) + (b_1(t))^2 \right) \, dt - \Phi_{0,t}^{-1} b_1(t) \, dW_t \right) \\
&\quad - (b_1(t) X_t + b_2(t)) \, dW_t \\
&= \Phi_{0,t}^{-1} (a_2(t) - b_1(t) b_2(t)) \, dt + \Phi_{0,t}^{-1} b_2(t) \, dW_t.
\end{align*}
\]

Integrating both sides, using \( \Phi_{0,0}^{-1} = 1 \), we obtain

\[ \Phi_{0,t}^{-1} X_t = X_0 + \int_0^t \Phi_{0,s}^{-1} (a_2(s) - b_1(s) b_2(s)) \, ds + \int_0^t \Phi_{0,s}^{-1} b_2(s) \, dW_s, \quad (4.23) \]

and hence we have established the following result.

**Proposition 4.2** The linear stochastic differential equation

\[ dX_t = (a_1(t) X_t + a_2(t)) \, dt + (b_1(t) X_t + b_2(t)) \, dW_t, \quad (4.24) \]

with initial value \( X_0 \) has the general solution

\[ X_t = \Phi_{0,t} \left( X_0 + \int_0^t \Phi_{0,s}^{-1} (a_2(s) - b_1(s) b_2(s)) \, ds + \int_0^t \Phi_{0,s}^{-1} b_2(s) \, dW_s \right), \quad (4.25) \]

where

\[ \Phi_{0,s} = \exp \left( \int_0^s \left( a_1(u) - \frac{1}{2} (b_1(u))^2 \right) du + \int_0^s b_1(u) \, dW_u \right), \quad (4.26) \]
Note that this solution reduces to the one given in (4.10) when $a_2(t) \equiv b_2(t) \equiv 0$, $a_1(t) = e(t)$ and $b_1(t) = f(t)$.

With the aid of the solution formula for (4.24) found in the Proposition above, we can also solve those nonlinear stochastic differential equations

$$dY_t = a(t, Y_t) \, dt + b(t, Y_t) \, dW_t,$$

that can be written as $X_t = U(t, Y_t)$, where $U$ is an (at least locally) invertible function and $X_t$ is linear, that is governed by an equation of the same form as (4.24).

Evaluating $X_t = U(t, Y_t)$ with Itô's formula gives

\[
\begin{align*}
dX_t &= \frac{\partial U}{\partial t}(t, Y_t) \, dt + \frac{\partial U}{\partial y}(t, Y_t) \, dY_t + \frac{1}{2} \frac{\partial^2 U}{\partial y^2}(t, Y_t) (dY_t)^2 \\
&= \left\{ \frac{\partial U}{\partial t}(t, Y_t) + \frac{\partial U}{\partial y}(t, Y_t) a(t, Y_t) + \frac{1}{2} \frac{\partial^2 U}{\partial y^2}(t, Y_t) (b(t, Y_t))^2 \right\} \, dt \\
&\quad + \frac{\partial U}{\partial y}(t, Y_t) b(t, Y_t) \, dW_t,
\end{align*}
\]

from which we can deduce the equivalences

\[
\frac{\partial U}{\partial t}(t, y) + \frac{\partial U}{\partial y}(t, y) a(t, y) + \frac{1}{2} \frac{\partial^2 U}{\partial y^2}(t, y) (b(t, y))^2 = a_1(t) U(t, y) + a_2(t),
\]

and

\[
\frac{\partial U}{\partial y}(t, y) b(t, y) = b_1(t) U(t, y) + b_2(t).
\]

This system of equations is so general that there is no hope of finding general criteria under which a solution can be found. Consequently we shall regard a special case (more general solutions are found in Section 4.3 of [KP92]).

**Example 4.3** The stochastic Verhulst model for the size $N_t$ of a population at time $t$, discussed in Section 1.1, is given by

\[
dN_t = \left( \lambda(t) N_t - N_t^2 \right) \, dt + \sigma(t) N_t \, dW_t,
\]

where we, for simplicity, have chosen the functions $\lambda(t)$ and $\sigma(t)$ to be constant. This is a non-linear equation, which can be written in the form of (4.27) if we set $a(t, y) = \lambda y - y^2$ and $b(t, y) = \sigma y$. We shall derive an expression for the solution $N_t$.

We hope to be able to relate $N_t$ to the solution $X_t$ of a linear stochastic differential equation

\[
dx_t = (a_1 x_t + a_2) \, dt + (b_1 x_t + b_2) \, dW_t.
\]

Set $X_t = U(N_t)$. Then, with $y = N_t$, (4.29)-(4.30) reduce to

\[
U'(N_t) \left( \lambda N_t - N_t^2 \right) + \frac{1}{2} U''(N_t) (\sigma N_t)^2 = a_1 U(N_t) + a_2,
\]

and

\[
U'(N_t) \sigma N_t = b_1 U(N_t) + b_2.
\]

If we assume that $b_1 \neq 0$ and $\sigma \neq 0$, (4.34) is separable differential equation with the general solution

\[
U(N_t) = CN_t^{\frac{b_1}{2}} - \frac{b_2}{b_1},
\]

37
where $C$ is an arbitrary constant. Substituting this expression for $U(N_t)$ into (4.33), we obtain

$$\frac{b_1}{\sigma} C N_t^\frac{\lambda}{\sigma} - 1 (\lambda N_t - N_t^2) + \frac{1}{2} \frac{b_1}{\sigma} \left( \frac{b_1}{\sigma} - 1 \right) C N_t^\frac{b_2}{\sigma} (\sigma N_t)^2 = a_1 \left( C N_t^\frac{b_2}{\sigma} - \frac{b_2}{b_1} \right) + a_2,$$

which is equal to

$$C N_t^\frac{\lambda}{\sigma} \left( b_1 \left( \frac{\lambda - N_t}{\sigma} - \frac{\sigma}{2} \right) + \frac{1}{2} \frac{b_1}{\sigma} - a_1 \right) = a_2 - a_1 \frac{b_2}{b_1}.$$

Since the right hand side is independent of $N_t$ we must determine which values on $b_1$ and $a_1$ that makes the left hand side independent of $N_t$ as well. Differentiating (4.37), multiplying the result with $C$ and differentiating once more yields the simple equation

$$\frac{b_1}{\sigma} \left( 1 + \frac{b_1}{\sigma} \right) = 0,$$

with the obvious solution $b_1 = -\sigma$ (as $b_1 \neq 0$). Inserting $b_1 = -\sigma$ into (4.37) we get

$$C \left( 1 + \frac{1}{N_t} (-\lambda + \sigma^2 - a_1) \right) = a_2 + \frac{a_1 b_2}{\sigma},$$

which shows that we must choose $a_1 = \sigma^2 - \lambda$. We can choose the remaining constants arbitrarily. One simple choice is $a_2 = C = 1$ and $b_2 = 0$.

Conclusively we get that $N_t$ is related to $X_t$ governed by

$$dX_t = ((\sigma^2 - \lambda) X_t + 1) \, dt - \sigma X_t \, dW_t,$$

through the relation

$$X_t = U(N_t) = N_t^{-1}.$$

For the linear stochastic differential equation in (4.40) we have the solution (see Proposition 4.2)

$$X_t = \Phi_{0,t} \left( X_0 + \int_0^t \Phi_{s,t}^{-1} \, ds \right),$$

where

$$\Phi_{0,s} = \exp \left( \int_0^s \left( \sigma^2 - \lambda - \frac{1}{2} (-\sigma)^2 \right) \, du + \int_0^s -\sigma \, dW_u \right) = \exp \left( - \left( \left( \lambda - \frac{1}{2} \sigma^2 \right) s + \sigma W_s \right) \right),$$

so the stochastic Verhulst model has the solution

$$N_t = \frac{1}{X_t} \frac{1}{\Phi_{0,t} \left( X_0 + \int_0^t \Phi_{0,s}^{-1} \, ds \right)} = \frac{\Phi_{0,t}^{-1}}{X_0 + \int_0^t \Phi_{0,s}^{-1} \, ds} = \frac{1}{N_0 + \int_0^t \exp \left( \frac{\lambda - \frac{1}{2} \sigma^2}{\sigma^2} s + \sigma W_s \right) \, ds}.$$

### 4.2 Existence and uniqueness of solutions

Existence and uniqueness of solutions to stochastic differential equations are shown in a similar fashion as for ordinary differential equations. We shall use the Picard-Lindelöf method of successive approximations of the solution and show that the limit of these approximations is the sought solution.
Theorem 4.4 (4.5.3 in [KP92]) Consider the stochastic differential equation

\[ X_t = X_0 + \int_0^t a(s, X_s) \, ds + \int_0^t b(s, X_s) \, dW_s, \tag{4.45} \]

where the random variable \( X_0 \) and the functions \( a \) and \( b \) satisfy the following conditions

1. The functions \( a(s,x) \) and \( b(s,x) \) are jointly \( \mathcal{L} \times \mathcal{L} \)-measurable on \([0,T] \times \mathbb{R}\).
2. There exists a positive constant \( K \) such that for all \( t \in [0,T] \) and \( x, y \in \mathbb{R} \)
   
   (a) \( |a(t,x) - a(t,y)| + |b(t,x) - b(t,y)| \leq K |x - y|, \)  \( \text{(Lipschitz continuity)} \)
   
   (b) \( |a(t,x)|^2 + |b(t,x)|^2 \leq K^2 \left( 1 + |x|^2 \right). \)  \( \text{(Linear growth)} \)
3. \( X_0 \) satisfies \( E[X_0^2] < \infty \) and is, for all \( t > 0 \), independent of the \( \sigma \)-algebra \( \mathcal{A}_t \) generated by \( W \).

Then there exists a solution \( X_t \) of (4.45) defined on \([0,T]\) which is continuous w.p.1, adapted with respect to the filtration \( \mathcal{F}_t \), generated by \( X_0 \) and \( W_s \) for \( s \leq t \), and satisfies

\[ \sup_{0 \leq t \leq T} E[X_t^2] < \infty. \tag{4.46} \]

Furthermore the solution is pathwise unique, that is if \( X \) and \( Y \) are two such solutions then

\[ P \left( \sup_{0 \leq t \leq T} |X_t - Y_t| = 0 \right) = 1. \tag{4.47} \]

Before proving this theorem, we consider two simple examples of stochastic differential equations (or actually ordinary differential equations since \( b(t,x) = 0 \) in both cases) violating conditions 2 (a) and 2 (b) respectively and confirm that existence or uniqueness of solutions fails.

Example 4.5 The equation

\[ \frac{dX_t}{dt} = X_t^2, \quad X_0 = 1, \tag{4.48} \]

corresponding to \( a(t,x) = x^2 \) and \( b(t,x) = 0 \) violates the condition of linear growth. This ordinary differential equation is separable and has the unique solution

\[ X_t = \frac{1}{1-t}, \text{ for } 0 \leq t < 1. \tag{4.49} \]

The solution explodes at \( t = 1 \) and hence the differential equation is not solvable on all finite time interval.

Example 4.6 The equation

\[ \frac{dX_t}{dt} = 3X_t^{2/3}, \quad X_0 = 0, \tag{4.50} \]

corresponding to \( a(t,x) = 3x^{2/3} \) and \( b(t,x) = 0 \) violates the Lipschitz continuity condition at \( x = 0 \) where \( \frac{\partial a}{\partial x} \) is infinite. This ordinary differential equation is separable and has the general solution \( X_t = (t + C)^3 \). One can show that for any \( a > 0 \) the function

\[ X_t = \begin{cases} 0, & \text{for } t \leq a \\ (t-a)^3, & \text{for } t > a \end{cases} \tag{4.51} \]

solves (4.50), so the solution is not unique.
Before proving Theorem 4.4 we state two lemmas, the first of which is left unproved (for a proof see [Eva98]).

**Lemma 4.7 (Gronwall inequality)** Let \( \alpha, \beta : [t_0, T] \to \mathbb{R} \) be nonnegative integrable functions such that
\[
\alpha (t) \leq \beta (t) + L \int_{t_0}^{t} \alpha (s) \, ds,
\]
for \( t \in [t_0, T] \) where \( L > 0 \). Then, for \( t \in [t_0, T] \),
\[
\alpha (t) \leq \beta (t) + L \int_{t_0}^{t} \exp (L (t - s)) \beta (s) \, ds.
\]

**Lemma 4.8** Let \( a \) and \( b \) satisfy conditions 1 and 2 in Theorem 4.4. Suppose that \( Y_i(t^i), i = 1, 2 \), is a stochastic process on \([0, T]\) such that
1. The increments \( W_{t+r} - W_t \), for \( r > 0 \) are independent of the \( \sigma \)-algebra generated by \( Y_i(t^i) \) and \( W_s \) for \( s \leq t \).
2. \( \sup_{0 \leq t \leq T} E \left[ \left( Y_i(t^i) \right)^2 \right] < \infty. \)

Let the process \( Z_i(t), t \in [0, T] \), be defined by
\[
Z_i(t) = X_0 + \int_0^{t} a \left( s, Y_i(s) \right) \, ds + \int_0^{t} b \left( s, Y_i(s) \right) \, dW_s.
\]

Then there exists a constant \( L > 0 \) such that
\[
E \left[ \left( Z_1(t) - Z_2(t) \right)^2 \right] \leq L \int_0^{t} E \left[ \left( Y_1(s) - Y_2(s) \right)^2 \right] \, ds.
\]

**Proof.** From the inequality \((a + b)^2 \leq 2a^2 + 2b^2\), we get the estimate
\[
\left| Z_1(t) - Z_2(t) \right|^2 = \left( \int_0^{t} \left( a \left( s, Y_1(s) \right) - a \left( s, Y_2(s) \right) \right) \, ds + \int_0^{t} \left( b \left( s, Y_1(s) \right) - b \left( s, Y_2(s) \right) \right) \, dW_s \right)^2
\leq 2 \left( \int_0^{t} \left( a \left( s, Y_1(s) \right) - a \left( s, Y_2(s) \right) \right) \, ds \right)^2 + 2 \left( \int_0^{t} \left( b \left( s, Y_1(s) \right) - b \left( s, Y_2(s) \right) \right) \, dW_s \right)^2.
\]

Using the following argument based on the Cauchy-Schwarz inequality
\[
\left( \int_0^{t} h(s) \, ds \right)^2 \leq \left( \int_0^{t} 1^2 \right) \left( \int_0^{t} (h(s))^2 \, ds \right) = t \int_0^{t} (h(s))^2 \, ds,
\]
together with the Lipschitz condition on the first term in (4.56), we obtain
\[
2 \left( \int_0^{t} \left( a \left( s, Y_1(s) \right) - a \left( s, Y_2(s) \right) \right) \, ds \right)^2 \leq 2t \int_0^{t} \left( a \left( s, Y_1(s) \right) - a \left( s, Y_2(s) \right) \right)^2 \, ds
\leq 2K^2 t \int_0^{t} \left( Y_1(s) - Y_2(s) \right)^2 \, ds.
\]
The Lipschitz condition and the second condition of the lemma assert that
\[ E \left[ \int_0^t \left( b \left( s, Y_s^{(1)} \right) - b \left( s, Y_s^{(2)} \right) \right)^2 ds \right] \leq K^2 E \left[ \int_0^t \left( Y_s^{(1)} - Y_s^{(2)} \right)^2 ds \right] < \infty \quad (4.59) \]
so \( b \left( s, Y_s^{(1)} \right) - b \left( s, Y_s^{(2)} \right) \in \mathcal{L}_T^2 \). Therefore by the Itô isometry
\[
E \left[ \left( \int_0^t \left( b \left( s, Y_s^{(1)} \right) - b \left( s, Y_s^{(2)} \right) \right) dW_s \right)^2 \right] = E \left[ \int_0^t \left( b \left( s, Y_s^{(1)} \right) - b \left( s, Y_s^{(2)} \right) \right)^2 ds \right] \leq 2K^2 \int_0^t E \left( Y_s^{(1)} - Y_s^{(2)} \right)^2 ds. \quad (4.60)
\]
Taking expectations of both sides of (4.56) and applying (4.58) and (4.60), we conclude the lemma follows by choosing
\[ L = 2K^2t + 2K^2. \quad (4.61) \]

**Proof of Theorem 4.4.** Define a sequence of approximations \( \left\{ X_t^{(n)} \right\}_{n=0}^\infty \) by \( X_t^{(0)} = X_0 \) and
\[
X_t^{(n)} = X_0 + \int_0^t a \left( s, X_s^{(n-1)} \right) ds + \int_0^t b \left( s, X_s^{(n-1)} \right) dW_s, \quad \text{for } n \geq 1. \quad (4.62)
\]
By induction and Theorem 3.12, the processes \( \left\{ X_t^{(n)} \right\}_{n=0}^\infty \) have continuous sample paths w.p.1. We shall show that the limit \( \tilde{X}_t = \lim_{n \to \infty} X_t^{(n)} \) is the unique solution, whose existence we are to verify. The proof consists of five steps, and we prove the theorem step by step.

(i) The sequence \( \left\{ X_t^{(n)} \right\} \) is uniformly mean-square bounded on \([0, T]\).

(ii) \( \left\{ X_t^{(n)} \right\} \) is uniformly mean square convergent.

(iii) \( \left\{ X_t^{(n)} \right\} \) is uniformly convergent w.p.1.

(iv) \( \tilde{X}_t = \lim_{n \to \infty} X_t^{(n)} \) is a solution to (4.45) and satisfies the properties of Theorem 4.4.

(v) Any two solutions \( X \) and \( Y \) of (4.45) agree w.p.1.

**Proof of (i).** We need to show that for some constant \( M \) and all positive integers \( n \)
\[
\sup_{0 \leq t \leq T} E \left[ \left( X_t^{(n)} \right)^2 \right] \leq M < \infty. \quad (4.63)
\]
We show this by induction over \( n \), where the first step obviously follows from condition 3. An application of the inequality \((a + b + c)^2 \leq 3a^2 + 3b^2 + 3c^2\) gives
\[
E \left[ \left( X_t^{(n)} \right)^2 \right] \leq 3E \left[ X_0^2 \right] + 3E \left[ \left( \int_0^t a \left( s, X_s^{(n-1)} \right) ds \right)^2 \right] + 3E \left[ \left( \int_0^t b \left( s, X_s^{(n-1)} \right) dW_s \right)^2 \right]. \quad (4.64)
\]
By the Cauchy-Schwarz inequality and Itô isometry, we obtain
\[
E \left[ (X_t^{(n)})^2 \right] \leq 3E \left[ X_0^2 \right] + 3TE \left[ \int_0^t \left( a \left( s, X_s^{(n-1)} \right) \right)^2 ds \right] + 3E \left[ \int_0^t \left( b \left( s, X_s^{(n-1)} \right) \right)^2 ds \right].
\]
(4.65)
Taking the linear growth condition into consideration and interchanging expectations and integrations, we get
\[
E \left[ (X_t^{(n)})^2 \right] \leq 3E \left[ X_0^2 \right] + 3K^2T \int_0^t \left( 1 + E \left[ (X_s^{(n-1)})^2 \right] \right) ds + 3K^2 \int_0^t \left( 1 + E \left[ (X_s^{(n-1)})^2 \right] \right) ds
\]
\[
\leq 3 \left\{ E \left[ X_0^2 \right] + L \int_0^t \left( 1 + E \left[ (X_s^{(n-1)})^2 \right] \right) ds \right\}
\]
\[
= 3 \left\{ E \left[ X_0^2 \right] + Lt + L \int_0^t E \left[ (X_s^{(n-1)})^2 \right] ds \right\},
\]
(4.66)
where the existence of the right hand side follows by the induction hypothesis stating that 
\[
\sup_{0 \leq t \leq T} E \left[ (X_t^{(n-1)})^2 \right] < \infty.
\]
Iterating this last inequality yields
\[
E \left[ (X_t^{(n)})^2 \right] \leq 3 \left\{ E \left[ X_0^2 \right] + Lt + L \int_0^t \left( 1 + E \left[ (X_s^{(n-2)})^2 \right] \right) ds \right\}
\]
\[
= 3 \left\{ (1 + 3Lt) E \left[ X_0^2 \right] + Lt + 3L^2 t^2 / 2 + L \int_0^t \int_s^t E \left[ (X_u^{(n-2)})^2 \right] du ds \right\}
\]
(4.67)
after one iteration and using the equality
\[
\int_0^t (t - s)^n \int_0^s f(r) \, dr \, ds = \int_0^t \frac{(t - s)^{n+1}}{(n + 1)} f(s) \, ds,
\]
(4.68)
we can show that after \(n\) iterations we get
\[
E \left[ (X_t^{(n)})^2 \right] \leq 3 \left\{ 1 + 3Lt + \ldots + \frac{(3Lt)^{n-1}}{(n-1)!} \right\} E \left[ X_0^2 \right] + 3Lt + \ldots + \frac{(3L)^n}{n!}
\]
\[
+ \left( 3L \right)^n \int_0^t \frac{(t - s)^{n-1}}{(n - 1)!} E \left[ X_0^2 \right] ds.
\]
(4.69)
Comparing this result with Taylor expansions of the exponential function it is clear that
\[
E \left[ (X_t^{(n)})^2 \right] \leq (3E \left[ X_0^2 \right] + 1) \exp (3LT) \leq M < \infty,
\]
(4.70)
so step (i) is verified.

**Proof of (ii).** We apply Lemma 4.8 to a sequence of successive approximations defined by
\[
(4.62)
\]
to obtain
\[
E \left[ (X_t^{(n+1)} - X_t^{(n)})^2 \right] \leq L \int_0^t E \left[ (X_s^{(n)} - X_s^{(n-1)})^2 \right] ds.
\]
(4.71)
Iterating, using (4.68), we get

$$E \left[ \left( X_t^{(n+1)} - X_t^{(n)} \right)^2 \right] \leq L^n \int_0^t (t-s)^{n-1} (n-1)! E \left[ \left( X_s^{(1)} - X_s^{(0)} \right)^2 \right] ds. \tag{4.72}$$

Applying the linear growth condition to the factor $E \left[ \left( X_s^{(1)} - X_s^{(0)} \right)^2 \right]$, we get

$$E \left[ \left( X_s^{(1)} - X_s^{(0)} \right)^2 \right] = E \left[ \left( \int_0^s a \left( u, X_u^{(0)} \right) du + \int_0^s b \left( u, X_u^{(0)} \right) dW_u \right)^2 \right]$$

$$\leq L \int_0^s (1 + E \left[ X_0^2 \right]) du \leq LT \left( 1 + E \left[ X_0^2 \right] \right) = C. \tag{4.73}$$

Calculating the integral in (4.72), we obtain

$$E \left[ \left( X_t^{(n+1)} - X_t^{(n)} \right)^2 \right] \leq \frac{C (LT)^n}{n!} \to 0, \tag{4.74}$$

as $n \to \infty$. The upper bound in (4.74) is independent of $t$, implying uniform mean square convergence of $\{ X_t^{(n)} \}$.

**Proof of (iii).** We now want to show uniform convergence w.p.1, that is

$$P \left( \omega \in \Omega : \left| X_t^{(n+1)} - X_t^{(n)} \right| \to 0 \text{ uniformly as } n \to \infty \right) = 1. \tag{4.75}$$

Let

$$Y^{(n)} = \sup_{0 \leq t \leq T} \left| X_t^{(n+1)} - X_t^{(n)} \right| \leq \sup_{0 \leq t \leq T} \left| \int_0^t \left( a \left( s, X_s^{(n)} \right) - a \left( s, X_s^{(n-1)} \right) \right) ds \right|$$

$$+ \sup_{0 \leq t \leq T} \left| \int_0^t \left( b \left( s, X_s^{(n)} \right) - b \left( s, X_s^{(n-1)} \right) \right) dW_s \right| = Y_1^{(n)} + Y_2^{(n)}. \tag{4.76}$$

We are interested in the expectation of the square of each of the terms $Y_i^{(n)}$. For $Y_1^{(n)}$ we apply the Cauchy-Schwarz inequality and the Lipschitz condition to obtain

$$E \left[ \left( Y_1^{(n)} \right)^2 \right] \leq E \left[ \left( \int_0^T \left( a \left( s, X_s^{(n)} \right) - a \left( s, X_s^{(n-1)} \right) \right) ds \right)^2 \right]$$

$$\leq T \int_0^T E \left[ \left( a \left( s, X_s^{(n)} \right) - a \left( s, X_s^{(n-1)} \right) \right)^2 \right] ds$$

$$\leq K^2 T \int_0^T E \left[ \left( X_s^{(n)} - X_s^{(n-1)} \right)^2 \right] ds, \tag{4.77}$$

and for $Y_2^{(n)}$ we apply the Doob inequality (Proposition 3.11) and the Lipschitz condition to obtain

$$E \left[ \left( Y_2^{(n)} \right)^2 \right] \leq E \left[ \left( \sup_{0 \leq t \leq T} \left| \int_0^t \left( b \left( s, X_s^{(n)} \right) - b \left( s, X_s^{(n-1)} \right) \right) dW_s \right| \right)^2 \right]$$

$$\leq 4E \left[ \int_0^T \left( b \left( s, X_s^{(n)} \right) - b \left( s, X_s^{(n-1)} \right) \right)^2 ds \right]$$

$$\leq 4K^2 \int_0^T E \left[ \left( X_s^{(n)} - X_s^{(n-1)} \right)^2 \right] ds. \tag{4.78}$$
Combining (4.76)-(4.78) and the inequality \((a + b)^2 \leq 2a^2 + 2b^2\) we get
\[
E \left[ \left( Y^{(n)} \right)^2 \right] \leq 2E \left[ \left( Y^{(n)}_1 \right)^2 \right] + 2E \left[ \left( Y^{(n)}_2 \right)^2 \right] \leq L \int_0^T E \left[ \left( X^{(n)}_s - X^{(n-1)}_s \right)^2 \right] ds
\]
\[
\leq \frac{C'(LT)^{n-1}}{(n-1)!},
\tag{4.79}
\]
where (4.74) was used in the last step. By the martingale inequality (Proposition 3.10) it follows from (4.79) that
\[
P \left\{ \frac{Y^{(n)}}{n^2} > \frac{1}{n^2} \right\} \leq \frac{C'(LT)^{n-1}}{(n-1)!} n^4.
\tag{4.80}
\]
The series
\[
\sum_{n=1}^{\infty} P \left\{ Y^{(n)} > \frac{1}{n^2} \right\},
\tag{4.81}
\]
must converge since it is majorized by a convergent series. By the Borel-Cantelli lemma then
\[
P \left\{ \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} Y^{(n)} > \frac{1}{n^2} \right\} = 0
\]
\[
\Leftrightarrow P \left\{ \sup_{0 \leq t \leq T} \left| X^{(n+1)}_t - X^{(n)}_t \right| \leq \frac{1}{n^2} \text{ for all } n \geq n_0 \text{ for some } n_0 \right\} = 1,
\tag{4.82}
\]
which proves the w.p.1 uniform convergence of \(\left\{ X^{(n)}_t \right\} \).

**Proof of (iv).** We investigate the properties of the limit process \(X_t = \lim_{n \to \infty} X^{(n)}_t\). Being the limit of \(A_t\)-measurable processes, \(X_t\) must be \(A_t\)-measurable as well. Furthermore, since \(X_t\) is the uniform limit of w.p.1 continuous processes, \(X_t\) itself must be continuous w.p.1. The mean-square boundedness of \(X_t\) (4.46) follows from the uniform mean square bound proved in step (i). Finally the uniform convergence and Lipschitz conditions imply
\[
\left| \int_0^t a \left( s, X^{(n)}_s \right) - a \left( s, \tilde{X}_s \right) \right| ds \leq K \int_0^t \left| X^{(n)}_s - \tilde{X}_s \right| ds \to 0,
\tag{4.83}
\]
and
\[
\int_0^t \left| b \left( s, X^{(n)}_s \right) - b \left( s, \tilde{X}_s \right) \right|^2 ds \leq K^2 \int_0^t \left| X^{(n)}_s - \tilde{X}_s \right|^2 ds \to 0,
\tag{4.84}
\]
w.p.1, which imply that
\[
\int_0^t a \left( s, X^{(n)}_s \right) ds \to \int_0^t a \left( s, \tilde{X}_s \right) \right) ds,
\tag{4.85}
\]
w.p.1 and, with the aid of the Itô isometry,
\[
\int_0^t b \left( s, X^{(n)}_s \right) dW_s \to \int_0^t b \left( s, \tilde{X}_s \right) dW_s,
\tag{4.86}
\]
in probability.

**Proof of (v).** Finally we establish uniqueness. Suppose that both \(X\) and \(Y\) are continuous w.p.1 and mean square bounded solutions to (4.45) on \([0, T]\). We use Lemma 4.8 on the function \(X_t - Y_t\), which must be bounded and measurable, and thus satisfies
\[
E \left[ (X_t - Y_t)^2 \right] \leq L \int_0^t E \left[ (X_s - Y_s)^2 \right] ds.
\tag{4.87}
\]
Gronwall’s inequality now asserts that $E\left[ (X_t - Y_t)^2 \right] = 0$. By the theorem of dominated convergence, we can conclude that $P(X_t = Y_t) = 1$ for any $t \in [0, T]$. If we take a countable dense subset $B$ of $[0, T]$ it still holds that

$$P(X_t = Y_t : t \in B) = 1.$$  

(4.88)

Now both $X_t$ and $Y_t$ are continuous w.p.1, so it must hold that

$$P\left( \sup_{0 \leq t \leq T} |X_t - Y_t| > 0 \right) = 0,$$  

(4.89)

which shows pathwise uniqueness. ■

4.3 Strong and weak solutions

Under some conditions, Theorem 4.4 asserted the existence of what we shall call a strong solution to a stochastic differential equation. By strong we mean that the version of the Wiener process was given in advance. Hence a strong solution is valid for any version of the Wiener process. It is shown in Theorem 4.6.1 in [KP92] that the strong solution $X_t$ to

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

(4.90)

asserted by Theorem 4.4 is actually a diffusion process with drift coefficient $a(t, X_t)$ and diffusion coefficient $b(t, X_t)$.

A question that naturally arises is if this result can be reversed, that is if any diffusion process is the solution to some stochastic differential equation. In Theorem 4.7.1 in [KP92] it is shown that, under some conditions on the drift and diffusion coefficients, there exists a weak solution to the stochastic differential equation corresponding to the diffusion process. A weak solution is not valid for any Wiener process but only for a particular choice of the Wiener process. Hence a weak solution consists of a pair $(X_t, \tilde{W}_t)$, where $\tilde{W}_t$ is the Wiener process for which the solution solves the given stochastic differential equation.

In Theorem 4.4 we showed pathwise uniqueness of strong solutions. Weak solutions, on the other hand, are only weakly unique, meaning that any two weak solutions to a given stochastic differential equation have the same distribution. We give an example of a stochastic differential equation that has only weak solutions and no strong solution.

**Example 4.9** Consider the stochastic differential equation

$$dX_t = \text{sgn}(X_t) \, dt + dW_t,$$  

(4.91)

with initial value $X_0 = 0$. Here

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ -1, & \text{if } x < 0. \end{cases}$$  

(4.92)

Let $X_t$ be a solution to this stochastic differential equation for some Wiener process $\tilde{W}_t$. Then $-X_t$ is a solution for the Wiener process $-\tilde{W}_t$. The two solutions $(X_t, \tilde{W}_t)$ and $(-X_t, -\tilde{W}_t)$ have the same distribution, but they are not pathwise unique, so no unique strong solution can exist.
5 Connecting stochastic and partial differential equations

5.1 Itô diffusions

In this section we consider $d$-dimensional stochastic differentials on the form

$$dX_t^{(i)} = a^{(i)}(X_t) \, dt + \sum_{k=1}^{m} b^{(i,k)}(X_t) \, dW_t^{(k)};$$

for $i = 1, \ldots, d$. For these, so-called autonomous stochastic differentials, the drift and diffusion coefficients have no direct time-dependence. In addition, we suppose that the prerequisites of Theorem 4.4 are satisfied. In the previous section we noted that, under these assumptions, stochastic differentials are diffusion processes. Hence, stochastic differentials on the form (5.1) will be referred to as Itô diffusions. It should also be noted that the absence of direct time-dependence is not crucial for the results in this section, but introducing assumption simplifies the proofs considerably.

We introduce some notation. Let $X_t^{s,x}$ denote the solution to (5.1) with the initial condition $X_s = x$. Furthermore let $\mathcal{M}$ denote the $\sigma$-algebra generated by the random variables $\omega \to X_t^{s,x}(\omega)$ and define the probability measure $Q^{s,x}$ on $\mathcal{M}$ by

$$Q^{s,x}(X_t^{L_1, \ldots, L_n} \in L_1, \ldots, X_t^{s,x} \in L_n),$$

where $L_i \in L$ and $t_i \in [s,T]$, for $1 \leq i \leq n$. Here $P$ is the probability measure induced by the solution to (5.1) with initial condition $X_s = 0$. By $E^{s,x}$ we mean expectation with respect to the probability measure $Q^{s,x}$. If $s = 0$, we write $X_t^x$, $Q^x$ and $E^x$ for $X_t^{0,x}$, $Q^{0,x}$ and $E^{0,x}$ respectively.

Itô diffusions satisfy the Markov property, that is the future behaviour of the process given what has happened up to time $t$ is the same as the behaviour obtained if we start the process at $X_t$. The precise mathematical formulation of this is the following:

**Theorem 5.1 (Markov property)** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function and let $X_t$ be the $d$-dimensional Itô diffusion (5.1). Then, for $t, h \geq 0$, it holds that

$$E^x \left[ f(X_{t+h}) | \mathcal{A}_t^{(m)} \right] = E^{X_t} \left[ f(X_h) \right].$$

The proof is elaborate and is left out (but can be found in Theorem 7.1.2 in [Øks98]). Itô diffusions also satisfy the Markov property for a more general time concept known as stopping times.

**Definition 5.2** Let $\{N_t\}$ be a filtration on a probability space $(\Omega, \mathcal{A}, P)$. A non-negative random variable $\tau$ is called a stopping time with respect to $\{N_t\}$ if the event

$$\{\omega \in \Omega : \tau(\omega) \leq t\} \in N_t,$$

for all $t \geq 0$.

This means that the information supplied by $N_t$ should always make it possible to decide whether $\tau(\omega) \leq t$ has occurred or not. Note that if $\tau(\omega) = t_0$ for all $\omega$, that is if $\tau$ is a deterministic time, then trivially $\tau$ is a stopping time with respect to any filtration, because in this case

$$\{\omega \in \Omega : \tau(\omega) \leq t\} = \begin{cases} \Omega, & \text{if } t \geq t_0, \\ \emptyset, & \text{if } t < t_0. \end{cases}$$

Two important subclasses of random times are the first exit or first hitting time of a set.
Theorem 5.4 (Strong Markov property) Let \( U \subset \mathbb{R}^d \) be an open set. Then the first exit time
\[
\tau_U = \inf \{ t > 0 : X_t \notin U \},
\]
is a stopping time with respect to \( \mathcal{M}_t \) (the filtration generated by \( X_t \)), since
\[
\{ \omega \in \Omega : \tau_U (\omega) \leq t \} = \bigcap_{m} \bigcup_{r \in \mathbb{Q} \cap (0, t]} \{ \omega \in \Omega : X_r \notin K_m \} \in \mathcal{M}_t,
\]
for all \( t \geq 0 \), where \( \{ K_m \} \) is an increasing sequence of closed sets such that \( U = \bigcup_m K_m \).

The first hitting time is defined similarly. We can define a \( \sigma \)-algebra \( \mathcal{M}_\tau \) with stopping time index as the \( \sigma \)-algebra generated by the sets \( \{ X_s \wedge \tau : s \geq 0 \} \); here \( s \wedge \tau \) is an abbreviation for \( \min (s, \tau) \).

We are now ready to state the strong Markov property. The proof is analogous to the proof of the Markov property and can be found in Theorem 7.2.4 in [Øks98].

Theorem 5.4 (Strong Markov property) Let \( f : \mathbb{R}^d \to \mathbb{R} \) be a bounded measurable function and let \( X_t \) be the \( d \)-dimensional Itô diffusion \( X_t \) (5.1). Let \( \tau \) be a stopping time with respect to \( \mathcal{A}_{\tau}^{(m)} \) such that \( \tau < \infty \) w.p.1. Then, for \( h \geq 0 \), it holds that
\[
E^x \left[ f (X_{\tau+h}) \mid \mathcal{A}_{\tau}^{(m)} \right] = E^{X_{\tau}} \left[ f (X_h) \right].
\]

The strong Markov property for Itô diffusions will be crucial when we, in the next section, derive the partial differential equation corresponding to a given Itô diffusion. As a first step in this direction we investigate the generator of a Itô diffusion.

Definition 5.5 Let \( X_t \) be the \( d \)-dimensional Itô diffusion (5.1). The (infinitesimal) generator \( A \) of \( X_t \) is defined by
\[
Af (x) = \lim_{t \to 0} \frac{E^x [f (X_t)] - f (x)}{t},
\]
for \( x \in \mathbb{R}^d \). The set of functions \( f : \mathbb{R}^d \to \mathbb{R} \) such that this limit exists at \( x \) is denoted \( D_A (x) \), while \( D_A \) denotes the set for which the limit exists for all \( x \in \mathbb{R}^d \).

In the following Lemma we use Itô’s formula to find an expansion for \( E^x [f (X_t)] \), which we subsequently use to relate the generator \( A \) to the drift and diffusion coefficients of (5.1).

Lemma 5.6 Let \( X_t^x \) be a \( d \)-dimensional stochastic differential defined by
\[
(X_t^x)^{(i)} = x^{(i)} + \int_0^t a^{(i)} (s, X_s^x) \, ds + \sum_{k=1}^m \int_0^t b^{(i,k)} (s, X_s^x) \, dW_s^{(k)},
\]
for \( i = 1, \ldots, d \) and some \( m \)-dimensional Wiener process \( W \). Let \( f \in C^2_0 (\mathbb{R}^d) \) and let \( \tau \) be a stopping time with respect to \( \mathcal{A}_{\tau}^{(m)} \) such that \( E^x [\tau] < \infty \). Assume that \( a (t, \omega) \) and \( b (t, \omega) \) are bounded for all \( (t, \omega) \) such that \( X_t^x \) belongs to the support of \( f \). Then
\[
E^x [f (X_{\tau})] = f (x) + E^x \left[ \int_0^\tau \left( \sum_{i=1}^d \frac{\partial f}{\partial x_i} (X_s) \, \frac{\partial a^{(i)} (s, X_s)}{\partial x_i} (X_s) \right) \, ds \right] + E^x \left[ \int_0^\tau \left( \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j} (X_s) \, \frac{\partial b^{(i,j)} (s, X_s)}{\partial x_i} (X_s) \right) \, ds \right],
\]
for \( \tau \geq 0 \) and \( x \in \mathbb{R}^d \).
where $E^x$ is defined as for Itô diffusions above. Note that (5.10) allow for direct time-dependence in the drift and diffusion coefficients of $X_t$.

**Proof.** Put $Y_t = f (X_t)$ (to simplify the notation we suppress the evaluation points) and apply Itô’s formula, see equation (3.90), to obtain

$$f (X_t) = f (X_0) + \int_0^t \left( \sum_{i=1}^d a^{(i)} (\partial f / \partial x_i) + \frac{1}{2} \sum_{i,j=1}^d b^{(i,k)} b^{(j,k)} (\partial^2 f / \partial x_i \partial x_j) \right) ds + \int_0^t \sum_{i=1}^d \sum_{k=1}^m b^{(i,k)} \partial f / \partial x_i dW_s^{(k)}. \quad (5.12)$$

Using the identity

$$\sum_{k=1}^m b^{(i,k)} b^{(j,k)} = \sum_{k=1}^m b^{(i,k)} (b^T)^{(k,j)} = (bb^T)^{(i,j)}, \quad (5.13)$$

replacing $t$ by $\tau$ and taking expectations on both sides of (5.12), noting that $E^x [f (X_0)] = f (x)$, the lemma follows immediately if we can show that

$$E^x \left[ \int_0^\tau \sum_{i=1}^d \sum_{k=1}^m b^{(i,k)} \partial f / \partial x_i dW_s^{(k)} \right] = 0. \quad (5.14)$$

Indeed, if $g$ is a bounded measurable function and $|g| \leq M$, then for all integers $k$ we have

$$E^x \left[ \int_0^{\tau \wedge k} g (X_s) \, dW_s \right] = E^x \left[ \int_0^{k \wedge s \wedge \tau} g (X_s) \, dW_s \right] = 0, \quad (5.15)$$

since $g (X_s)$ and $\chi_{\{s < \tau\}}$ are both $A_s$-measurable. It remains to show that

$$E^x \left[ \int_0^{\tau \wedge k} g (X_s) \, dW_s \right] = \lim_{k \to \infty} E^x \left[ \int_0^{\tau \wedge k} g (X_s) \, dW_s \right] = 0, \quad (5.16)$$

In Chapter 1.2 in [GS79] it is shown that the Itô isometry holds also when the interval of integration has stopping time limits. Hence

$$E^x \left[ \left( \int_0^\tau g (X_s) \, dW_s - \int_0^{\tau \wedge k} g (X_s) \, dW_s \right)^2 \right] = E^x \left[ \left( \int_{\tau \wedge k}^\tau g (X_s) \, dW_s \right)^2 \right]
= E^x \left[ \int_{\tau \wedge k}^\tau (g (X_s))^2 \, ds \right]
\leq M^2 E^x \left[ \int_{\tau \wedge k}^\tau ds \right] = M^2 E^x [\tau - \tau \wedge k] \quad (5.17)$$

which tends to zero as $k \to \infty$ since $E^x [\tau] < \infty$. Finally, (5.14) follows by setting $g = \sum_{i=1}^d b^{(i,k)} \partial f / \partial x_i$.

**Theorem 5.7** Let $X_t$ be the $d$-dimensional Itô diffusion (5.1). If $f \in C^2_b (\mathbb{R}^d)$, then $f \in DA$ and

$$Af (x) = \sum_{i=1}^d a^{(i)} (x) \partial f / \partial x_i + \frac{1}{2} \sum_{i,j=1}^d (bb^T)^{(i,j)} (x) \partial^2 f / \partial x_i \partial x_j. \quad (5.18)$$

**Proof.** This result follows directly from the definition of $A$ and Lemma 5.6 with $\tau = t$. ■

48
In many applications we are interested in the generator of a process where \( f \in C^2_0(\mathbb{R} \times \mathbb{R}^d) \). In this case we regard the auxiliary system of stochastic differential equations

\[
\begin{bmatrix}
    dY^{(0)}_s \\
    dY^{(1)}_s \\
    \vdots \\
    dY^{(d)}_s
\end{bmatrix} = 
\begin{bmatrix}
    a^{(1)} \left( Y^{(1)}_s, \ldots, Y^{(d)}_s \right) \\
    a^{(1)} \left( Y^{(1)}_s, \ldots, Y^{(d)}_s \right) \\
    \vdots \\
    a^{(1)} \left( Y^{(1)}_s, \ldots, Y^{(d)}_s \right)
\end{bmatrix} ds + 
\begin{bmatrix}
    b^{(1)} \left( Y^{(1)}_s, \ldots, Y^{(d)}_s \right) \\
    b^{(1)} \left( Y^{(1)}_s, \ldots, Y^{(d)}_s \right) \\
    \vdots \\
    b^{(1)} \left( Y^{(1)}_s, \ldots, Y^{(d)}_s \right)
\end{bmatrix} dW_s,
\]

with initial value \((Y^{(0)}_0, Y^{(1)}_0, \ldots, Y^{(d)}_0) = (t, x_1, \ldots, x_d) = (t, x) := y\). The generator of \( Y_s = (Y^{(0)}_s, Y^{(1)}_s, \ldots, Y^{(d)}_s) \) is given by

\[
Af(t, x) = \sum_{i=0}^d a^{(i)}(y) \frac{\partial f}{\partial y_i} + \frac{1}{2} \sum_{i,j=0}^d (bb^T)^{(i,j)}(y) \frac{\partial^2 f}{\partial y_i \partial y_j}
\]

\[
= \frac{\partial f}{\partial t} + \sum_{i=1}^d a^{(i)}(x) \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d (bb^T)^{(i,j)}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}. \tag{5.20}
\]

**Example 5.8** The \( d \)-dimensional Wiener process is an Itô diffusion with \( a^{(i)} = 0 \) and \( b^{(i,j)} = \delta_{ij} \), where \( \delta_{ij} \) is the Kronecker delta function. Hence for \( f \in C^2_0(\mathbb{R}^d) \), the generator of the Wiener process is

\[
Af(x) = \frac{1}{2} \sum_{i,j=1}^d \delta_{ik} \delta_{jk} \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{1}{2} \sum_{i=1}^d \frac{\partial^2 f}{\partial x_i^2} = \frac{1}{2} \Delta f, \tag{5.21}
\]

where \( \Delta \) is the Laplace operator.

A simple consequence of Lemma 5.6 and Theorem 5.7 is Dynkin’s formula

**Theorem 5.9 (Dynkin’s formula)** Let \( f \in C^2_0(\mathbb{R}^d) \) and \( X_t \) be the \( d \)-dimensional Itô diffusion (5.1). Assume that \( \tau \) is a stopping time such that \( E^x[\tau] < \infty \). Then

\[
E^x[f(X_{\tau})] = f(x) + E^x \left[ \int_0^\tau Af(X_s) \, ds \right]. \tag{5.22}
\]

As a detour, we can use Dynkin’s formula to prove some interesting properties of the multi-dimensional Wiener process.

**Example 5.10** Assume that \( W = (W^{(1)}, \ldots, W^{(d)}) \) is a \( d \)-dimensional Wiener process starting at \( a = (a_1, \ldots, a_d) \in \mathbb{R}^d \) with \( |a| < R \) for some positive constant \( R \). Determine the expected value of the first exit time \( \tau_R \) from the ball

\[
K_R = \left\{ x \in \mathbb{R}^d : |x| < R \right\}. \tag{5.23}
\]

Choose an arbitrary integer \( k \) and apply Dynkin’s formula with the finite stopping time \( \tau_R \land k \) and some \( C^2_0 \)-function \( f \) such that \( f(|x|^2) \) on \( K_R \).

\[
E^a[f(W_{\tau_R \land k})] = f(a) + E^a \left[ \int_0^{\tau_R \land k} Af(W_s) \, ds \right] = |a|^2 + E^a \left[ \int_0^{\tau_R \land k} \frac{1}{2} \Delta(|x|^2) \, ds \right]
\]

\[
= |a|^2 + dE^a[\tau_R \land k]. \tag{5.24}
\]
as \( \frac{1}{2} \Delta (|x|^2) = d \). Since \( E^a [f(W_{\tau R \wedge k})] \leq E^a [f(W_{\tau R})] = R^2 \) it is clear that

\[
E^a [\tau_R \wedge k] \leq \frac{1}{d} \left( R^2 - |a|^2 \right),
\]

(5.25) for all \( k \). Since \( E^a [\tau_R \wedge k] \) is uniformly bounded, it follows that

\[
E^a [\tau_R] = \lim_{k \to \infty} E^a [\tau_R \wedge k] < \infty.
\]

(5.26) Furthermore, by the theorem of dominated convergence, we obtain

\[
E^a [\tau_R] = \frac{1}{d} \left( R^2 - |a|^2 \right).
\]

(5.27) Hence the Wiener process exits any finite ball in finite time w.p.1.

**Example 5.11** Assume that \( d \geq 2 \) and \( b = (b_1, \ldots, b_d) \in \mathbb{R}^d \) with \( |b| > R \) for some positive constant \( R \). We would like to determine the probability that the Wiener process starting at \( b \) hits the sphere

\[
K_R = \left\{ x \in \mathbb{R}^d : |x| < R \right\}.
\]

(5.28) If this occurs with probability one we say that the Wiener process is recurrent and otherwise we call it transient.

Let \( \alpha_k \) be the first exit time from the annulus

\[
A_k = \left\{ x \in \mathbb{R}^d : R < |x| < 2^k R \right\},
\]

(5.29) for \( k = 1, 2, \ldots \). By the previous example we know that \( E^b [\alpha_k] < \infty \). Furthermore let \( T_R \) be the first hitting time of \( K_R \) and let \( f = f_{d,k} \) be a \( C_0^2 (\mathbb{R}^d) \)-function satisfying

\[
f (x) = \begin{cases} 
- \log |x|, & \text{for } d = 2, \\
|x|^{2-d}, & \text{for } d > 2,
\end{cases}
\]

(5.30) on \( A_k \). By construction \( f \) is the fundamental solution to the Laplace operator on \( A_k \) so on the annulus we have \( \Delta f = 0 \). Hence by Dynkin’s formula we get (note that \( E^b [\alpha_k] < \infty \) is required here)

\[
E^b [f (W_{\alpha_k})] = f (b),
\]

(5.31) for all \( k \). The exit from \( A_k \) can occur at the inner or at the outer boundary so it is natural to define the two probabilities

\[
p_k = P^b (|W_{\alpha_k}| = R) \quad \text{and} \quad q_k = P^b (|W_{\alpha_k}| = 2^k R).
\]

(5.32) We now consider the two cases \( d = 2 \) and \( d > 2 \) separately starting with the former case.

For \( d = 2 \), plugging the fundamental solution (5.30) into (5.31) we get

\[
p_k (- \log R) + q_k \left( - \log 2^k R \right) = - \log |b|,
\]

(5.33) for all \( k \). This implies that \( q_k \to 0 \) as \( k \to \infty \). But as \( p_k + q_k = 1 \), we must have that \( p_k \to 1 \) as \( k \to \infty \). We conclude that

\[
P^b (T_R < \infty) = \lim_{k \to \infty} p_k = 1,
\]

(5.34)
so the Wiener process is recurrent for $d = 2$. For $d > 2$, we get the equation

$$p_k R^{2-d} + q_k \left(2^k R\right)^{2-d} = |b|^{2-d}. \quad (5.35)$$

In this case $q_k$ does not vanish (since $(2^k R)^{2-d} \to 0$ as $k \to \infty$) and we obtain

$$P^b (T_R < \infty) = \lim_{k \to \infty} p_k = \left(\frac{|b|}{R}\right)^{2-d}, \quad (5.36)$$

so the Wiener process is transient for $d > 2$.

### 5.2 The Kolmogorov equations and Feynman-Kac’s formula

Let $X_t$ be the $d$-dimensional Itô diffusion (5.1). By setting $\tau = t$ and $\tau = t + h$, respectively, in Dynkin’s formula the following calculation shows that for $f \in C^2_0(\mathbb{R}^d)$ the function

$$u(t, x) = E^x [f(X_t)], \quad (5.37)$$

is differentiable with respect to $t$

$$\frac{\partial u}{\partial t} = \lim_{h \to 0} \frac{E^x [f(X_{t+h})] - E^x [f(X_t)]}{h} = \lim_{h \to 0} \frac{\left(f(x) + E^x \left[\int_0^h A f(X_s) \, ds\right]\right) - \left(f(x) + E^x \left[\int_0^h A f(X_s) \, ds\right]\right) h}{h} = \lim_{h \to 0} \frac{E^x \left[\int_0^h A f(X_s) \, ds\right]}{h} = E^x [Af(X_t)]. \quad (5.38)$$

It turns out that the right hand side of (5.38) can be expressed in terms of $u$ as well.

**Theorem 5.12 (Kolmogorov’s backward equation)** Let $f \in C^2_0(\mathbb{R}^d)$ and define $u(t, x)$ as in (5.37). Then $u(t, \cdot) \in D_A$ for all $t$ and

$$\begin{cases}
\frac{\partial u}{\partial t} = Au, & \text{for } t > 0, x \in \mathbb{R}^d, \\
u(0, x) = f(x), & \text{for } x \in \mathbb{R}^d.
\end{cases} \quad (5.39)$$

Moreover if $w(t, x) \in C^{1,2} (\mathbb{R}^+ \times \mathbb{R}^d)$ is a bounded solution to (5.39), then $w(t, x) = u(t, x)$ with $u$ as in (5.37).

**Proof.** Fix $t$ in $u(t, x)$. Then, since $t \to u(t, x)$ is differentiable, the Markov property and the first statement of Proposition 1.17 implies

$$\frac{E^x [u(t, X_h)] - u(t, x)}{h} = \frac{E^x [E^{X_t} [f(X_t)]] - u(t, x)}{h} = \frac{E^x [E^{X_t} [f(X_{t+h})] | A_h]] - u(t, x)}{h} = \frac{E^x [f(X_{t+h})] - u(t, x)}{h} = \frac{u(t+h, x) - u(t, x)}{h} \to \frac{\partial u}{\partial t}, \quad (5.40)$$

as $h \to 0$. Hence

$$Au = \lim_{h \to 0} \frac{E^x [u(t, X_h)] - u(t, x)}{h}, \quad (5.41)$$

exists and (5.39) follows.

51
For the converse statement, assume that the function \( w(t, x) \in C^{1,2}(\mathbb{R} \times \mathbb{R}^d) \) satisfies (5.39). Then
\[
\bar{A}w := \frac{\partial w}{\partial t} + Aw = 0, \quad \text{for } t > 0, \ x \in \mathbb{R}^d
\]
and \( w(0, x) = f(x) \) for \( x \in \mathbb{R}^d \). Fix \( (t, x) \in \mathbb{R} \times \mathbb{R}^d \) and define the process \( Y_s \in \mathbb{R}^{d+1} \) as \( Y_s = \left( t - s, X_s^0, x \right) \) for \( s \geq 0 \). Here \( X_s^0, x \) represents the solution to (5.1) started at \( x \) at time \( 0 \), so \( (t, x) \) is the initial point of \( Y_s \). By (5.20), it is clear that \( \bar{A} \) is the generator of \( Y_t \) and according to Dynkin’s formula we have
\[
E^{t,x}[w(Y_{\tau_R})] = w(t, x) + \int_0^{\tau_R} \bar{A}w(Y_s) \, ds = w(t, x),
\]
where \( \tau_R = \inf \{ s > 0 : |X_s| \geq R \} \). Letting \( R \to \infty \), (5.43) reduces to \( w(t, x) = E^{t,x}[w(Y_s)] \) for all \( s \geq 0 \). Finally setting \( s = t \), the initial condition of (5.39) implies
\[
w(t, x) = E^{t,x}[w(Y_t)] = E \left[ w \left( 0, X_t^0, x \right) \right] = E \left[ f \left( X_t^0, x \right) \right] = E^x[f(X_t)],
\]
which completes the proof. \( \blacksquare \)

We next consider the derivation of the related Kolmogorov’s forward equation. Assume that the transition probability of the Itô diffusion \( X_t \) has a smooth density \( p(s, x; t, y) \), so that
\[
E^x[f(X_t)] = \int_{\mathbb{R}^d} f(y) p(0, x; t, y) \, dy,
\]
for \( f \in C^2_0(\mathbb{R}^d) \). By Dynkin’s formula with \( \tau = t \) we get
\[
\int_{\mathbb{R}^d} f(y) p(0, x; t, y) \, dy = E^x[f(X_t)] = f(x) + E^x \left[ \int_0^t A f(X_s) \, ds \right]
= f(x) + \int_0^t E^x[A f(X_s)] \, ds = f(x) + \int_0^t \int_{\mathbb{R}^d} (A f(y)) p(0, x; s, y) \, dy \, ds.
\]
For test functions \( \phi, \psi \in C^2_0(\mathbb{R}^d) \), we use integration by parts to obtain the relation
\[
(A \phi, \psi) = \int_{\mathbb{R}^d} (A \phi)(y) \psi(y) \, dy = \int_{\mathbb{R}^d} \left( \sum_{i=1}^d a^{(i)}(y) \frac{\partial \phi}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^d (b_i b_j^T)^{(i,j)}(y) \frac{\partial^2 \phi}{\partial y_i \partial y_j} \right) \psi(y) \, dy
= \int_{\mathbb{R}^d} \phi(y) \left( -\sum_{i=1}^d \frac{\partial}{\partial y_i} \left( a^{(i)}(y) \psi(y) \right) + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial y_i \partial y_j} \left( (b_i b_j^T)^{(i,j)}(y) \psi(y) \right) \right) \, dy
= \int_{\mathbb{R}^d} \phi(y) A^* \psi(y) \, dy = (\phi, A^* \psi),
\]
so if we differentiate (5.46) with respect to \( t \) and apply (5.47), we arrive at
\[
\int_{\mathbb{R}^d} f(y) \frac{\partial p(0, x; t, y)}{\partial t} \, dy = \int_{\mathbb{R}^d} (A f(y)) p(0, x; t, y) \, dy = \int_{\mathbb{R}^d} f(y) A^* p(0, x; t, y) \, dy,
\]
(5.48)
which is a weak formulation of the partial differential equation
\[ \frac{\partial p}{\partial t} = A^* p, \text{ for } t > 0, x \in \mathbb{R}^d, \] (5.49)
known as Kolmogorov’s forward equation or the Fokker-Planck equation. In this case the initial condition is
\[ \lim_{s \to t} p(s, x; t, y) = \delta(x - y), \text{ for } x \in \mathbb{R}^d, \] (5.50)
where \( \delta \) is Dirac’s delta function.

Observe that we can reverse the time and transform Kolmogorov’s backward equation from an initial value problem to a terminal value problem. Introduce the function \( \tilde{u} \) as
\[ \tilde{u}(t, x) = u(T - t, x) = E^x [f(X_{T-t})] = E^{t,x} [f(X_T)], \] (5.51)
where the last step follows by the time-homogenity of Itô diffusions (a proof of this property can be found in Section 7.1 of [Øks98]). Then \( \tilde{u} \) solves the terminal value problem
\[ \begin{cases} \frac{\partial \tilde{u}}{\partial t} + A\tilde{u} = 0, & \text{for } t \in [0, T], x \in \mathbb{R}^d, \\ \tilde{u}(T, x) = f(x), & \text{for } x \in \mathbb{R}^d. \end{cases} \] (5.52)

The prefixes forward and backward originate from the fact that in the backward equation in terminal value form we differentiate with respect to the starting time and location of the process whereas, in the forward equation, we differentiate with respect to the future time and location of the process. The Kolmogorov equations connect Itô diffusions with drift and diffusion coefficients \( a \) and \( b \) to partial differential equations with coefficients \( a \) and \( b \).

We shall now prove a useful generalization of Kolmogorov’s backward equation, known as the Feynman-Kac formula.

**Theorem 5.13 (Feynman-Kac’s formula)** Let \( f \in C^2_0 (\mathbb{R}^d) \), \( q \in C(\mathbb{R}^d) \) and assume that \( q \) is bounded from below. Put
\[ u(t, x) = E^x \left[ f(X_t) \exp \left( - \int_0^t q(X_r) \, dr \right) \right], \] (5.53)
where \( X_t \) is the \( d \)-dimensional Itô diffusion (5.1). Then
\[ \begin{cases} \frac{\partial u}{\partial t} = Au - qu, & \text{for } t > 0, x \in \mathbb{R}^d, \\ u(0, x) = f(x), & \text{for } x \in \mathbb{R}^d. \end{cases} \] (5.54)

Moreover if \( w(t, x) \in C^{1,2}(\mathbb{R} \times \mathbb{R}^d) \) is bounded on \( K \times \mathbb{R}^d \) for every compact \( K \subset \mathbb{R} \) and \( w \) solves (5.54), then \( w(t, x) = u(t, x) \) with \( u \) as in (5.53).

**Proof.** The proof is analogous to, but somewhat more complicated than, the proof of Kolmogorov’s backward equation. Define the stochastic differentials \( Y_t \) and \( Z_t \) as
\[ Y_t = f(X_t), \quad Z_t = \exp \left( - \int_0^t q(X_r) \, dr \right), \] (5.55)
so that
\[ E^x [Y_t Z_t] = E^x \left[ f(X_t) \exp \left( - \int_0^t q(X_r) \, dr \right) \right] = u(t, x). \] (5.56)
With this definition $dY_t$ is given by (5.12) and $dZ_t$ by

$$dZ_t = -q(X_t) Z_t \, dt,$$

(5.57)

and, since $dY_t \, dZ_t = 0$ (because $dZ_t$ contains no $dW_t$-term),

$$d(Y_t Z_t) = Y_t \, dZ_t + Z_t \, dY_t.$$

(5.58)

From (5.58) and the definitions of $Y_t$ and $Z_t$ it follows that $Y_t Z_t$ is a Itô diffusion. Hence, by (5.56) and Lemma 5.6, $u(t, x)$ is differentiable with respect to $t$. We rewrite $E^x [u(t, X_h)]$ as

$$E^x [u(t, X_h)] = E^x \left[ E^{X_h} \left[ f(X_t) \exp \left( -\int_0^t q(X_r) \, dr \right) \right] \right],$$

(5.59)

so in analogy to (5.40), we get

$$\lim_{h \to 0} \frac{E^x [u(t, X_h)] - u(t, x)}{h} = \lim_{h \to 0} \frac{E^x \left[ f(X_{t+h}) Z_{t+h} \exp \left( \int_0^h q(X_r) \, dr \right) \right] - u(t, x)}{h}
= \lim_{h \to 0} \frac{E^x \left[ f(X_{t+h}) Z_{t+h} \right] - u(t, x)}{h}
+ \lim_{h \to 0} \frac{E^x \left[ f(X_{t+h}) Z_{t+h} \left( \exp \left( \int_0^h q(X_r) \, dr \right) - 1 \right) \right]}{h}
= \frac{\partial u}{\partial t}(t, x) + q(x) u(t, x),$$

(5.60)

The last limit follows from the theorem of dominated convergence since

$$\frac{f(X_{t+h}) Z_{t+h} \left( \exp \left( \int_0^h q(X_r) \, dr \right) - 1 \right)}{h} \to f(X_t) Z_t q(X_0)$$

pointwise and boundedly as $h \to 0$.

For the converse statement assume that the function $w(t, x) \in C^{1,2}(\mathbb{R} \times \mathbb{R}^d)$ satisfies (5.54) and that $w(t, x)$ is bounded on $K \times \mathbb{R}^d$ for every compact $K \subset \mathbb{R}$. Then

$$\tilde{A} w := -\frac{\partial w}{\partial t} + A w - q w = 0,$$

(5.62)

for $t > 0, x \in \mathbb{R}^d$ and $w(0, x) = f(x)$ for $x \in \mathbb{R}^d$. Fix $(t, x, z) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$ and define the processes $Z$ and $H$ as

$$Z_s = z + \int_0^s q(X_r) \, dr,$$

(5.63)

and

$$H_s = (t - s, X_0^{0,x}, Z_s),$$

(5.64)

54
respectively. Note that \((t,x,z)\) is the initial point of \(H\). It follows from the definition of \(H\) that it is a Itô diffusion so by a generalization of \((5.20)\) it has the generator \(A_H\)
\[
A_H \phi(t,x,z) = -\frac{\partial \phi}{\partial t} + A \phi + q(x) \frac{\partial \phi}{\partial x},
\]
for \(\phi \in C^2_0(\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d)\). By Dynkin’s formula we have, for all \(t \geq 0\) and \(R \geq 0\)
\[
E^{t,x,z} [\phi(H_{s \wedge \tau_R})] = \phi(t,x,z) + E^{t,x,z} \left[ \int_0^{s \wedge \tau_R} A_H \phi(H_r) \, dr \right],
\]
where \(\tau_R = \inf \{ s > 0 : |H_s| \geq R \}\). If we now set \(\phi(t,x,z) = \exp(-z) w(t,x)\) and use \((5.54)\), we get
\[
A_H \phi(t,x,z) = \exp(-z) \left( \frac{\partial w}{\partial t} + Aw - q(x) w \right) = 0,
\]
so for this choice of \(\phi\)
\[
w(t,x) = \phi(t,x,0) = E^{t,x,0} [\phi(H_{s \wedge \tau_R})]
\]
\[
= E^x \left[ \exp \left( - \int_0^{s \wedge \tau_R} q(X_r) \, dr \right) w(t - (s \wedge \tau_R), X_{s \wedge \tau_R}) \right]
\]
\[
\rightarrow E^x \left[ \exp \left( - \int_0^s q(X_r) \, dr \right) w(t - s, X_s) \right],
\]
as \(R \to \infty\), since \(w(t,x)\) is bounded on \(K \times \mathbb{R}^d\) and \(q\) is lower bounded. Finally setting \(s = t\), we obtain, from \((5.54)\) and \((5.68)\),
\[
w(t,x) = E^x \left[ \exp \left( - \int_0^t q(X_r) \, dr \right) w(0, X_t) \right] = E^x \left[ \exp \left( - \int_0^t q(X_r) \, dr \right) f(X_t) \right]
\]
\[
= u(t,x),
\]
and the proof is finished. \(\blacksquare\)

5.3 The Black-Scholes equation and application to finance

In this section we investigate an application of the Feynman-Kac formula from the field of financial mathematics. As a start we derive the Black-Scholes equation for the price of a European call option already mentioned in Section 1.2.

Example 5.14 Let \(p(t,x)\) be the price of a European call option, where \(x = X_t\) is the price of a share satisfying the stochastic differential equation
\[
dX_t = \mu X_t \, dt + \sigma X_t \, dW_t,
\]
where the drift \(\mu\) and volatility \(\sigma\) are constant. Assume the existence of a risk-free paper \(B_t\) satisfying
\[
dB_t = r B_t \, dt,
\]
where \(r\) is a constant representing the risk free rent. Find a partial differential equation for \(p(t,x)\).
Consider the portfolio replicating the option
\[ I_t = -p(t, X_t) + \alpha(t) X_t + \beta(t) B_t, \]  
(5.72)
for \( \alpha(t), \beta(t) \in \mathbb{R} \). We assume that the portfolio is self-financing in the sense that no money is brought in or taken out of the portfolio. Mathematically this can be written
\[ d(\alpha(t) X_t + \beta(t) B_t) = \alpha(t) \, dX_t + \beta(t) \, dB_t, \]
(5.73)
for the portfolio \( \alpha(t) X_t + \beta(t) B_t \). Hence using Itô's formula on \( I_t \), we obtain
\[
\begin{aligned}
\frac{d}{dt} \left( \alpha(t) X_t + \beta(t) B_t \right) &= \alpha(t) \frac{dX_t}{dt} + \beta(t) \frac{dB_t}{dt}
\end{aligned}
\]
(5.74)
Now choose \( \beta(t) \) so that the portfolio becomes riskless (only the deterministic part remains). This can be achieved by setting \( \beta(t) = \frac{\partial p}{\partial x} \). Then
\[
\begin{aligned}
\frac{d}{dt} I_t &= -\frac{\partial p}{\partial t} dt - \frac{\partial p}{\partial x} dX_t - \frac{1}{2} \frac{\partial^2 p}{\partial x^2} (dX_t)^2 + \alpha(t) \, dX_t + \beta(t) \, dB_t \\
&= \left( -\frac{\partial p}{\partial t} \frac{\partial p}{\partial x} \mu X_t - \frac{1}{2} \frac{\partial^2 p}{\partial x^2} \sigma^2 X_t^2 + \alpha(t) \mu X_t + \beta(t) \sigma B_t \right) \, dt \\
&\quad + \left( \alpha(t) - \frac{\partial p}{\partial x} \right) \sigma X_t \, dW_t.
\end{aligned}
\]
(5.75)
If we now suppose that the opportunity of arbitrage is precluded, so that the probability of gaining money without risk is zero, then \( I_t \) must evolve as the risk-free paper \( B_t \). Hence \( dI_t = rI_t \, dt \), so in comparison to (5.75), we get
\[
\begin{aligned}
\left( -\frac{\partial p}{\partial t} \frac{\partial p}{\partial x} \mu X_t - \frac{1}{2} \frac{\partial^2 p}{\partial x^2} \sigma^2 X_t^2 + \beta(t) \sigma B_t \right) \, dt &= dI_t = rI_t \, dt = r \left( -p + \alpha(t) X_t + \beta(t) B_t \right) \, dt \\
&= r \left( -p + \frac{\partial p}{\partial x} X_t + \beta(t) B_t \right) \, dt.
\end{aligned}
\]
(5.76)
which, after rearranging the terms and setting \( x = X_t \), reduces to the partial differential equation
\[
\frac{\partial p}{\partial t} + r x \frac{\partial p}{\partial x} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 p}{\partial x^2} = r p,
\]
(5.77)
for \( t < T \), where \( T \) is the exercise time of the option. At time \( T \) the cost of exercising the option is \( K \) so the terminal condition corresponding to (5.77) must be
\[
p(T, x) = (x - K)^+. \]
(5.78)
Equation (5.77) continues to hold also when the risk free rent and volatility are allowed to depend on time and the share value. In these cases, however, we may not be able to solve the equation explicitly and we must use finite differential or finite element methods or, alternatively, the following representation supplied by the Feynman-Kac formula.
First observe that, in exactly the same manner as for Kolmogorov’s backward equation, the Feynman-Kac formula can be formulated as a terminal value problem by reversing the time variable. We introduce the function $\tilde{u}$ as

$$
\tilde{u}(t, x) = u(T - t, x) = E^x \left[ f(X_{T-t}) \exp \left( - \int_0^{T-t} q(X_r) \, dr \right) \right]
$$

$$
= E^x \left[ f(X_T) \exp \left( - \int_t^T q(X_r) \, dr \right) \right],
$$

(5.79)

where we have used the time-homogeneity of Itô diffusions in the last step. It then follows that $\tilde{u}$ solves the terminal value problem

$$
\frac{\partial \tilde{u}}{\partial t} + A \tilde{u} - q \tilde{u} = 0, \quad \text{for } t \in [0, T], \ x \in \mathbb{R},
$$

(5.80)

If we set $a(x) = rx$ and $b(x) = \sigma x$ in (5.1) and in addition set $q(x) = r$ and $f(x) = (x - K)^+$, the terminal value problem (5.80) is transformed into (5.77)-(5.78). Hence the solution to the Black-Scholes equation can be represented

$$
p(t, x) = \tilde{u}(t, x) = E^x \left[ f(X_T) \exp \left( - \int_t^T q(X_s) \, ds \right) \right]
$$

$$
= E^x \left[ (X_T - K)^+ \exp \left( - \int_t^T r \, ds \right) \right]
$$

$$
= E^x \left[ (X_T - K)^+ \exp \left( -r(T - t) \right) \right],
$$

(5.81)

as was previously stated in (1.10). Finally inserting the explicit solution to the geometric Brownian motion (5.70) derived in Example 4.1 into (5.81), we obtain

$$
p(t, x) = E \left[ x e^{(r - \frac{1}{2} \sigma^2)(T - t) + \sigma W_{T-t}} - K \right]^+ e^{-r(T - t)}
$$

$$
= \int_{\mathbb{R}} \left[ x e^{(r - \frac{1}{2} \sigma^2)(T - t) + \sigma y} - K \right]^+ e^{-r(T - t)} \frac{1}{\sqrt{2\pi(T - t)}} e^{-\frac{y^2}{2(T - t)}} \, dy.
$$

(5.82)

The Feynman-Kac formula can be shown to hold for more general Itô processes so this kind of representation continues to hold also in cases when the share price is given by more complex stochastic differential equations. In these cases we may not be able to solve (5.81) explicitly, but, nevertheless, we can use the numerical methods for stochastic differential equations described in subsequent Chapters to approximate the option price $p(t, x)$.

### 6 Stochastic optimal control

In this chapter we derive methods to solve problems of the kind studied in Section 1.3, where we discussed the situation in which a person has a portfolio $Y_t$ consisting of risky shares $X_t = \alpha(t) Y_t$ and riskless bonds $B_t = (1 - \alpha(t)) Y_t$, where $\alpha(t) \in [0, 1]$ is a control function and

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

(6.1)

$$
dB_t = c(t) B_t \, dt,
$$

(6.2)
with \( 0 \leq c(t) < a(t) \). For a given function \( g \) we wish to maximize the cost function
\[
C_{t,y}(\alpha) = E^{t,y}[g(Y_T)].
\] (6.3)

Using (6.1)-(6.2), we can derive a stochastic differential equation governing the evolution of the portfolio \( Y_t \)
\[
dY_t = d(\alpha(t) Y_t + (1 - \alpha(t)) Y_t) = dX_t + dB_t = a(t) X_t dt + b(t) X_t dW_t + c(t) B_t dt
\]
\[
= a(t) \alpha(t) Y_t dt + b(t) \alpha(t) Y_t dW_t + c(t) (1 - \alpha(t)) Y_t dt
\]
\[
= (a(t) \alpha(t) + c(t) (1 - \alpha(t))) Y_t dt + b(t) \alpha(t) Y_t dW_t.
\] (6.4)

Equation (6.4) is used in the following example, where the explicit form of \( g \) is known, and will be considered in further detail at the end of this chapter.

**Example 6.1** What is the optimal control function \( \alpha = \alpha^* \) for \( a(t) \), \( b(t) \) and \( c(t) \) constant and \( g(y) = y^r \), with \( 0 < r < 1 \)?

From (6.3) it is natural to investigate \( g(Y_t) = (Y_t)^r \). By Itô’s formula, we obtain
\[
d(g(Y_t)) = d((Y_t)^r) = rY_t^{r-1} dY_t + \frac{1}{2} r (r-1) Y_t^{r-2} (dY_t)^2
\]
\[
= rY_t^{r-1} (a\alpha + c (1 - \alpha) Y_t) dt + b\alpha Y_t dW_t + \frac{1}{2} r (r-1) Y_t^{r-2} b^2 \alpha^2 Y_t^2 dt.
\] (6.5)

Integrating and taking expectations, we obtain
\[
E^{t,y}[g(Y_T)] = g(y) + E^{t,y}\left[ \int_t^T d(g(Y_s)) \right] = y^r + E^{t,y}\left[ \int_t^T rY_s^{r-1} b\alpha dW_s \right]
\]
\[
+ E^{t,y}\left[ \int_t^T rY_s^{r-1} \left( a\alpha + c (1 - \alpha) + \frac{(r-1) b^2 \alpha^2}{2} \right) ds \right].
\] (6.6)

We perturb the above equation with respect to some \( \varepsilon \in \mathbb{R} \) sufficiently small for \( \alpha = \alpha^* + \varepsilon \in [0,1] \) to hold for some admissible control function \( v \). Then the optimal control \( \alpha^* \) must satisfy
\[
C_{t,y}(\alpha^* + \varepsilon) \leq C_{t,y}(\alpha^*),
\] (6.7)

for all \( v \). Applying the expansion in (6.6) in (6.7), we get
\[
C_{t,y}(\alpha^* + \varepsilon v) - C_{t,y}(\alpha^*)
\]
\[
= E^{t,y}\left[ \int_t^T rY_s^{r-1} \left( a(\alpha^* + \varepsilon v) + c (1 - (\alpha^* + \varepsilon v)) + \frac{(r-1) b^2 (\alpha^* + \varepsilon v)^2}{2} \right) ds \right]
\]
\[
- E^{t,y}\left[ \int_t^T rY_s^{r-1} \left( a\alpha^* + c (1 - \alpha^*) + \frac{(r-1) b^2 \alpha^*}{2} \right) ds \right]
\]
\[
= E^{t,y}\left[ \int_t^T rY_s^{r-1} \left( a\varepsilon v - c \varepsilon v + \frac{(r-1) b^2 ((\alpha^* + \varepsilon v)^2 - (\alpha^*)^2)}{2} \right) ds \right]
\]
\[
= E^{t,y}\left[ \int_t^T rY_s^{r-1} \left( a - c + (r-1) b^2 \alpha^* + \frac{\varepsilon v}{2} \right) ds \right].
\] (6.8)
Since the sign of \( \varepsilon v \) may vary and can be made arbitrarily small, it is clear that we can only assert that (6.7) holds if
\[
a - c + (r - 1) b^2 \alpha^* = 0,
\]  
that is if
\[
\alpha^* = \frac{a - c}{b^2 (1 - r)}.
\]

We can formulate the problem of stochastic optimal control in a more general setting as follows. Let \( X_s \) be the stochastic differential
\[
dX_s^{(i)} = a^{(i)}(s, X_s, \alpha(s, X_s)) \, ds + \sum_{k=1}^{m} b^{(i,j)}(s, X_s, \alpha(s, X_s)) \, dW_s^{(k)},
\]
for \( i = 1, \ldots, d, \, t \leq s \leq T \) with initial condition \( X_t = x \). Here \( T \) is a fixed terminal time and \( x \in \mathbb{R}^d \) a given initial point. Assume that \( a^{(i)}, b^{(i,j)} : [0, T] \times \mathbb{R}^d \times A \to \mathbb{R} \) are smooth, bounded functions, where \( A \) is a compact subset of \( \mathbb{R}^d \). We call the function \( \alpha : [0, T] \times \mathbb{R}^d \to A \) a control and let \( A \) be the set of admissible Markov controls \( s \to \alpha(s, X_s) \) (what we mean by admissible will vary from problem to problem). The Markov control uses the current values of \( s \) and \( X_s \) to affect the dynamics of the system by adjusting the drift and diffusion coefficients. For the admissible controls \( \alpha \in A \), we define the cost
\[
C_{t,x}(\alpha) = E_t^x \left[ \int_t^T h(s, X_s, \alpha(s, X_s)) \, ds + g(X_T) \right],
\]
where \( X_s \) solves the stochastic differential equation (6.11) and \( h : [0, T] \times \mathbb{R}^d \times A \to \mathbb{R} \) and \( g : \mathbb{R}^d \to \mathbb{R} \) are given smooth, bounded functions. We call \( h \) the running cost and \( g \) the terminal cost. Our goal is to find an optimal control \( \alpha^* \) which minimizes the expected cost \( C_{t,x}(\alpha) \).

Define the value function
\[
u(t, x) := \inf_{\alpha \in A} C_{t,x}(\alpha).
\]
We shall show that \( \nu \) solves a certain partial differential equation and that the optimal control \( \alpha^* \) can be reconstructed from \( \nu \). For simplicity we assume that the optimal control is attained, that is
\[
\nu(t, x) = \min_{\alpha \in A} C_{t,x}(\alpha) = C_{t,x}(\alpha^*).
\]
The following theorem holds also without this assumption, but the proof is more cumbersome in that case. Before proving the theorem we state and prove an auxiliary lemma.

**Theorem 6.2** Assume that \( X_t \) solves (6.11) with a Markov control function \( \alpha \) and suppose that the function \( \nu \) defined by (6.13) is bounded and smooth. Then \( \nu \) satisfies the Hamilton-Jacobi equation
\[
\min_{\alpha \in A} \{H(\alpha, t, x, u)\} = 0, \quad \text{for} \quad t \in [0, T], \, x \in \mathbb{R}^d,
\]
\[
\nu(T, x) = g(x), \quad \text{for} \quad x \in \mathbb{R}^d,
\]
where \( H \) is the Hamiltonian
\[
H(\alpha, t, x, u) = \frac{\partial u}{\partial t} + \sum_{i=1}^{d} a^{(i)}(t, x, \alpha) \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \left( b^T \right)^{(i,j)}(t, x, \alpha) \frac{\partial^2 u}{\partial x_i \partial x_j} + h(t, x, \alpha).
\]

59
Lemma 6.3 Assume that \( X_t, \alpha, g, h \) pose a general stochastic optimal control problem according to the assumptions made above. Then the function \( u \) satisfies, for all \( \delta > 0 \), the relation
\[
u(t, x) = \min_{\alpha: [t, t+\delta] \rightarrow A} E^{t,x} \left[ \int_t^{t+\delta} h(s, X_s, \alpha(s, X_s)) \, ds + u(t + \delta, X_{t+\delta}) \right]. \tag{6.17} \]

**Proof.** We first use the optimal control on \([t, T]\), denoted \(\alpha^*\), to verify that \(u\) is minorized by the right hand side of (6.17). Since \(\alpha^*\) is optimal, we have
\[
u(t, x) = \min_{\alpha \in A} E^{t,x} \left[ \int_t^T h(s, X_s, \alpha(s, X_s)) \, ds + g(X_T) \right] \tag{6.18} \]
\[
= E^{t,x} \left[ \int_t^{t+\delta} h(s, X_s, \alpha^*(s, X_s)) \, ds \right] + E^{t,x} \left[ \int_{t+\delta}^T h(s, X_s, \alpha^*(s, X_s)) \, ds + g(X_T) \right].
\]
By properties of the conditional expectation, we have
\[
E^{t,x} \left[ \int_{t+\delta}^T h(s, X_s, \alpha^*(s, X_s)) \, ds + g(X_T) \right] \geq E^{t,x} \left[ u(t + \delta, X_{t+\delta}) \right].
\tag{6.19}
\]
and hence
\[
u(t, x) \geq E^{t,x} \left[ \int_t^{t+\delta} h(s, X_s, \alpha^*(s, X_s)) \, ds \right] + E^{t,x} \left[ u(t + \delta, X_{t+\delta}) \right] \geq \min_{\alpha: [t, t+\delta] \rightarrow A} E^{t,x} \left[ \int_t^{t+\delta} h(s, X_s, \alpha(s, X_s)) \, ds + u(t + \delta, X_{t+\delta}) \right]. \tag{6.20}
\]

For the converse statement that \(u\) is majorized by the right hand side of (6.17), we choose the arbitrary control \(\alpha^+\) on \([t, t+\delta]\) and then, given the value of \(X_{t+\delta}\), choose the optimal control \(\alpha^*\) on \([t+\delta, T]\). Denote this control by \(\alpha' = (\alpha^+, \alpha^*)\). By the definition of \(u\) we get
\[
u(t, x) \leq C_{t,x}(\alpha') = E^{t,x} \left[ \int_t^T h(s, X_s, \alpha'(s, X_s)) \, ds + g(X_T) \right] \tag{6.21}
\]
Taking the minimum over all controls on \([t, t + \delta]\) now yields the sought inequality and the lemma follows. \( \blacksquare \)

**Proof of Theorem 6.2.** Choose \(\alpha = \alpha^*\) to be the optimal control for (6.11) on \([t, t + \delta]\). Then by Lemma 6.3
\[
u(t, X_t) = E^{t,X_t} \left[ \int_t^{t+\delta} h(s, X_s, \alpha^*(s, X_s)) \, ds + u(t + \delta, X_{t+\delta}) \right]. \tag{6.22}
\]
so an application of Itô’s theorem (compare to the proof of Lemma 5.6) gives

\[-E^{t,X_t} \left[ \int_t^{t+\delta} h(s, X_s, \alpha^*(s, X_s)) \, ds \right] = E^{t,X_t} \left[ u(t+\delta, X_{t+\delta}) - u(t, X_t) \right] \]

\[
= E^{t,X_t} \left[ \int_t^{t+\delta} \left( \frac{\partial u}{\partial t} + \sum_{i=1}^{d} a^+_i \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \sum_{k=1}^{m} b_{i,k}^b b_{j,k}^b \frac{\partial^2 u}{\partial x_i \partial x_j} \right) \, ds \right], \tag{6.23}
\]

where \( a_* = a(t, X_t, \alpha^*(t, X_t)) \) and \( b_* = b(t, X_t, \alpha^*(t, X_t)) \). Furthermore, by setting \( h_* = h(t, X_t, \alpha^*(t, X_t)) \) it follows directly from (6.23) that

\[
\frac{\partial u}{\partial t} + \sum_{i=1}^{d} a_*^{(i)} \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \sum_{k=1}^{m} b_{i,k}^b b_{j,k}^b \frac{\partial^2 u}{\partial x_i \partial x_j} + h_* = 0, \tag{6.24}
\]

since \( \sum_{k=1}^{m} b_{i,k}^b b_{j,k}^b = (b b^T)^{(i,j)} \). Moreover, we can conclude from (6.12)-(6.13) that

\[
u(T, x) = g(x). \tag{6.25}
\]

Now choose the an arbitrary control for (6.11) on \([t, t+\delta]\) and denote this choice by \( \alpha = \alpha^+ \). By the proof of Lemma 6.3 we obtain

\[
u(t, x) \leq E^{t,x} \left[ \int_t^{t+\delta} h(s, X_s, \alpha^+(s, X_s)) \, ds + u(t+\delta, X_{t+\delta}) \right], \tag{6.26}
\]

implying that

\[-E^{t,X_t} \left[ \int_t^{t+\delta} h(s, X_s, \alpha^+(s, X_s)) \, ds \right] \leq E^{t,X_t} \left[ u(t+\delta, X_{t+\delta}) - u(t, X_t) \right]. \tag{6.27}
\]

Following (6.23), we get the inequality

\[
\frac{\partial u}{\partial t} + \sum_{i=1}^{d} a_+^{(i)} \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \sum_{k=1}^{m} b_{i,k}^b b_{j,k}^b \frac{\partial^2 u}{\partial x_i \partial x_j} + h_+ \geq 0, \tag{6.28}
\]

where \( a_+ = a(t, X_t, \alpha^+(t, X_t)) \), \( b_+ = b(t, X_t, \alpha^+(t, X_t)) \) and \( h_+ = h(t, X_t, \alpha^+(t, X_t)) \). Combining equations (6.24), (6.25) and (6.28) the theorem follows. \( \blacksquare \)

Applying Theorem 6.2 to (6.4), it is clear that in the optimal portfolio example the maximal cost \( u(t, y) = \max_{\alpha \in A} E^{t,y} [g(Y_T)] \) will be the solution to the Hamilton-Jacobi equation

\[
\max_{\alpha \in A} \{ H(\alpha, t, y, u) \} = 0, \quad \text{for } t \in [0, T], \, y \in \mathbb{R}^d, \quad u(T, y) = g(y), \tag{6.29}
\]

with

\[
H(\alpha, t, y, u) = \frac{\partial u}{\partial t} + \sum_{i=1}^{d} a^{(i)}(t, y, \alpha) \frac{\partial u}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^{d} (bb^T)^{(i,j)}(t, y, \alpha) \frac{\partial^2 u}{\partial y_i \partial y_j} + h(t, y, \alpha)
\]

\[
= \frac{\partial u}{\partial t} + (a(t) \alpha(t) + c(t) (1 - \alpha(t))) y \frac{\partial u}{\partial y} + \frac{(b(t) \alpha(t))^2 y^2}{2} \frac{\partial^2 u}{\partial y^2}. \tag{6.30}
\]

61
Example 6.4 Assume that the functions $a(t)$, $b(t)$ and $c(t)$ in (6.4) are constants and that $rac{\partial^2 u}{\partial y^2} < 0$ in (6.30). Find the optimal control function $\alpha^*$ and the Hamilton-Jacobi equation of this problem?

Differentiating the function

$$f(\alpha) = \frac{\partial u}{\partial t} + (a\alpha + c(1-\alpha))y\frac{\partial u}{\partial y} + \frac{b^2\alpha^2y^2}{2}\frac{\partial^2 u}{\partial y^2}, \quad (6.31)$$

with respect to $\alpha$ yields

$$\frac{\partial f}{\partial \alpha} = (a-c)y \frac{\partial u}{\partial y} + b^2y^2 \frac{\partial^2 u}{\partial y^2}. \quad (6.32)$$

Setting this derivative equal to zero, we obtain

$$\hat{\alpha} = \frac{(c-a)\frac{\partial u}{\partial y}}{b^2y\frac{\partial^2 u}{\partial y^2}}. \quad (6.33)$$

The optimal control $\alpha^*$ is thus given by

$$\alpha^* = \begin{cases} 
0, & \text{for } \hat{\alpha} < 0, \\
\hat{\alpha}, & \text{for } \hat{\alpha} \in [0,1], \\
1, & \text{for } \hat{\alpha} > 0.
\end{cases} \quad (6.34)$$

The left hand side of the Hamilton-Jacobi equation can now easily be found to be

$$\max_{\alpha \in A} \{H(\alpha, t, x, u)\} = \begin{cases} 
\frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y}, & \text{for } \hat{\alpha} < 0, \\
\frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y} - \frac{(c-a)^2}{2}\left(\frac{\partial u}{\partial y}\right)^2, & \text{for } \hat{\alpha} \in [0,1], \\
\frac{\partial u}{\partial t} + ay\frac{\partial u}{\partial y} + \frac{b^2y^2}{2}\frac{\partial^2 u}{\partial y^2}, & \text{for } \hat{\alpha} > 0.
\end{cases} \quad (6.35)$$

If we solve (6.29) with $\max_{\alpha \in A} \{H(\alpha, t, y, u)\}$ as in Example 6.4, we get the maximal cost function for the optimal portfolio problem. Unfortunately this is quite a hard problem, but an approximate solution can be found by numerical methods for partial differential equations.

7 An introduction to numerical approximation

The remainder of this course is devoted to the study of numerical methods for stochastic differential equations. In this introductory chapter we define a number of concepts that are central to this field, such as different types of numerical error, strong and weak approximations and stability. We shall begin by defining the simplest numerical approximation for stochastic differential equations, the Euler approximation.
7.1 Euler approximation

Consider the stochastic differential

\[ dX_t = a(t, X_t) \, dt + b(t, X_t) \, dW_t, \quad (7.1) \]

on \( 0 \leq t \leq T \), with initial value \( X_0 \). Let \( 0 = t_0 < t_1 < \ldots < t_N = T \) be a partition of the interval \([0, T]\), from now on referred to as a time discretization of \([0, T]\). The Euler approximation of \( X_t \) is a continuous time stochastic process \( \{X_t \, : \, t \in [0, T]\} \) given by

\[ X_{t_{n+1}} = X_{t_n} + a(t_n, X_{t_n}) \, (t_{n+1} - t_n) + b(t_n, X_{t_n}) \, (W_{t_{n+1}} - W_{t_n}), \quad (7.2) \]

for \( n = 0, 1, \ldots, N - 1 \) with initial value \( X_{t_0} = X_0 \). We shall use the notation

\[ \Delta t_n = t_{n+1} - t_n, \quad (7.3) \]

for the \( n \)th time increment and

\[ \delta = \max_n \Delta t_n, \quad (7.4) \]

for the maximal time increment. In most cases considered below the time discretization will be equidistant, so that

\[ \delta = \Delta t_n = T/N, \quad (7.5) \]

for every \( n = 0, 1, 2, \ldots, N - 1 \).

The main difference compared to the Euler approximation for ordinary differential equations is that an increment \( W_{t_{n+1}} - W_{t_n} \) of the Wiener process has to be generated at every time step. These increments are Gaussian with mean zero and variance \( \Delta t_n \), so \( W_{t_{n+1}} - W_{t_n} \in N \left( 0, \Delta t_n \right) \) and we can, for example, use the Box-Müller algorithm described below to generate the Wiener process increments.

**Theorem 7.1 (Box-Müller method)** If \( U_1 \) and \( U_2 \) are two independent uniformly distributed random variables on \([0, 1]\), then \( N_1 \) and \( N_2 \) given by

\[ N_1 = \sqrt{-2 \log(U_2) \cos(2\pi U_1)}, \quad (7.6) \]

\[ N_2 = \sqrt{-2 \log(U_2) \sin(2\pi U_1)}, \quad (7.7) \]

are two independent Gaussian variables (that is \( N_1, N_2 \in N(0,1) \)).

**Proof.** Manipulating (7.6)-(7.7), we obtain

\[ U_1 = \frac{1}{2\pi} \arctan \left( \frac{N_2}{N_1} \right), \quad (7.8) \]

\[ U_2 = \exp \left( -\frac{N_1^2 + N_2^2}{2} \right), \quad (7.9) \]

The joint density of \( N_1 \) and \( N_2 \), \( p_{N_1, N_2}(n_1, n_2) \), is given by

\[ p_{N_1, N_2}(n_1, n_2) = p_{U_1, U_2}(u_1(n_1, n_2), u_2(n_1, n_2)) \frac{\partial(u_1, u_2)}{\partial(n_1, n_2)} \]

\[ = 1 \begin{vmatrix} \frac{1}{2\pi} & -n_2 \\ -n_1 \exp \left( -\frac{n_1^2 + n_2^2}{2} \right) & -n_2 \exp \left( -\frac{n_1^2 + n_2^2}{2} \right) \end{vmatrix} \]

\[ = \frac{1}{2\pi} \exp \left( -\frac{n_1^2 + n_2^2}{2} \right), \]

which equals the joint density of two independent \( N(0,1) \) variables.
Although the Box-Müller method is quite simple, the evaluation of trigonometric function makes it time consuming. An alternative method is the Polar-Marsaglia method considered in Section 1.3 of [KP92].

For a given time discretization, the Euler method determines values of the approximation at the discretization times only. If necessary we can use some interpolation method to find approximate values at intermediate times as well. The simplest choice of interpolation is the **piecewise constant interpolation** in which we define

\[ X_t = X_{t_n(t)}, \]

where \( n(t) = \max \{ n \in \mathbb{N} : t_n \leq t \} \). More frequently used, however, is the **linear interpolation**

\[ X_t = X_{t_n(t)} + \frac{t - t_n(t)}{t_{n(t)+1} - t_n(t)} \left( X_{t_{n(t)+1}} - X_{t_n(t)} \right), \]

which is continuous.

### 7.2 Strong and weak approximation

We need a criterion to judge the quality of different numerical approximation methods. It turns out that there are two types of tasks connected to the simulation of solutions to stochastic differential equations. The first occurs in the case when we are interested in a good pathwise approximation, for example if we are to simulate the trajectories of a solution. The second task involves the approximation of expectations of functions of the solution, such as the probability distribution or moments of the solution. These kinds of problems are of special importance as expectations of functions are often encountered in applications (compare for example with (5.81)) and can seldom be determined explicitly.

Let \( X \) be the explicit solution to a stochastic differential equation on the time interval \([0, T]\) and let \( 0 = t_0 < t_1 < \ldots < t_N = T \) be a time discretization of \([0, T]\). By a **time discrete approximation** we mean a stochastic process \( \overline{X} = \{ \overline{X}_t : t \in [0, T] \} \) which is based on the time discretization in the sense that \( \overline{X}_{t_n} \) is \( \mathcal{F}_{t_n} \)-measurable and \( \overline{X}_{t_{n+1}} \) can be expressed as a function of \( \overline{X}_{t_0}, \overline{X}_{t_1}, \ldots, \overline{X}_{t_n}, t_0, t_1, \ldots, t_n, t_{n+1} \) and a finite number of \( \mathcal{F}_{t_{n+1}} \)-measurable random variables. Every time discrete approximation corresponds to a **scheme** describing a recursive algorithm for the generation of values at the discretization points and an interpolation method used to generate values at intermediate points. Hence \( \overline{X}_{t_n} \) is constructed in order to provide an approximate value of \( X_{t_n} \) and the **absolute error** of the time discrete approximation is defined as

\[ \varepsilon = E \left[ \left| X_T - \overline{X}_T \right| \right]. \]

This gives an estimate of the pathwise closeness at the end of the discretization interval. We say that the time discrete approximation \( \overline{X} \) with maximum time increment \( \delta \) **converges strongly** to \( X \) at time \( T \) if

\[ \lim_{\delta \to 0} E \left[ \left| X_T - \overline{X}_T \right| \right] = 0. \]

Hence pathwise approximation is known as **strong approximation**.

In order to compare different methods of strong approximation, we say that the time discrete approximation \( \overline{X} \) converges strongly to \( X \) with **order** \( \gamma \) at time \( T \) if there exists a constant \( C \), independent of \( \delta \), and a constant \( \delta_0 > 0 \) such that

\[ E \left[ \left| X_T - \overline{X}_T \right| \right] \leq C \delta^\gamma, \]

64
for each \( \delta \in (0, \delta_0) \). In Theorem 7.4 below we prove that under some regularity conditions, the Euler approximation converges strongly with order \( \gamma = 0.5 \).

If, on the other hand, approximations of expectations of functions of \( X \) is requested, it is sufficient to consider the mean error

\[
\mu = |E[X_T] - E[\bar{X}_T]|,
\]

or more generally the mean error of some function \( g \)

\[
|E[g(X_T)] - E[g(\bar{X}_T)]|.
\]

This error estimates the proximity of the distributions of \( X_T \) and \( \bar{X}_T \). We say that the time discrete approximation \( \bar{X} \) with maximum step size \( \delta \) converges weakly to \( X \) at time \( T \) with respect to a class \( C \) of test functions \( g : \mathbb{R}^d \to \mathbb{R} \) if

\[
\lim_{\delta \to 0} |E[g(X_T) - g(\bar{X}_T)]| = 0,
\]

for all \( g \in C \). Hence approximation of functions of the solution is known as weak approximation.

The customary choice for the space of test functions is the space \( C^l_\ell (\mathbb{R}^d, \mathbb{R}) \) of \( l \) times continuously differentiable functions \( g : \mathbb{R}^d \to \mathbb{R} \), which, together with their partial derivatives up to and including order \( l \), have polynomial growth. This space obviously contains the space of all polynomials and will be sufficient for most purposes. With \( C \) specified as \( C^l_\ell (\mathbb{R}^d, \mathbb{R}) \), we say that the time discrete approximation \( \bar{X} \) converges weakly to \( X \) with order \( \beta \) at time \( T \) if there for each \( g \in C^l_\ell (\mathbb{R}^d, \mathbb{R}) \) exists a constant \( C \), independent of \( \delta \), and a constant \( \delta_0 > 0 \) such that

\[
|E[g(X_T) - g(\bar{X}_T)]| \leq C \delta^\beta,
\]

for each \( \delta \in (0, \delta_0) \). In Theorem 7.6 below we prove that the Euler approximation converges weakly with order \( \beta = 1 \).

### 7.3 Statistical error and time discretization error

In this section we investigate the various kinds of error that arise when approximating the solution to stochastic differential equations numerically. First, we consider the case where the stochastic differential equation can be solved explicitly, because in this case any numerical solution can be compared to the exact solution. One such example is the geometric Brownian motion \( X \) given by

\[
dX_t = aX_t \, dt + bX_t \, dW_t,
\]

with the solution

\[
X_t = X_0 \exp \left( \left( a - \frac{1}{2} b^2 \right) t + bW_t \right).
\]

Now let \( \bar{X} \) be a time discrete approximation of \( X \). If we are to compare the paths of \( X \) and \( \bar{X} \) we must use the same sample path of the Wiener process. Hence the values of the exact solution at the discretization times are given by

\[
X_{t_n} = X_0 \exp \left( \left( a - \frac{1}{2} b^2 \right) t_n + b \sum_{j=1}^{n} (W_{t_j} - W_{t_{j-1}}) \right),
\]

where \( W_{t_j} - W_{t_{j-1}} \) are the Wiener process increments used in the time discrete approximation.
We simulate $M$ trajectories of the Wiener process and use these trajectories to construct $M$ approximate solutions $\mathbf{X}$ and $M$ exact solutions $\mathbf{X}$. We let the final value of the $k$th approximate and exact solution be denoted by $\mathbf{X}_{T,k}$ and $\mathbf{X}_{T,k}$, respectively. Using the simulated trajectories, the absolute error defined in (7.13) can be estimated by

$$\hat{\varepsilon} = \frac{1}{M} \sum_{k=1}^{M} |X_{T,k} - X_{T,k}|.$$  \hfill (7.23)

By the Central limit theorem (see page 25 in [KP92]), we know that $\hat{\varepsilon}$ asymptotically tends to a Gaussian random variable and that it converges weakly (in the same sense as in the previous section) to the absolute error $\varepsilon$ as $M \to \infty$. Obviously, we cannot generate infinitely many trajectories. Nevertheless it is possible to estimate the variance $\sigma^2_{\hat{\varepsilon}}$ of $\hat{\varepsilon}$ and use this variance to construct a confidence interval for $\varepsilon$.

We arrange the simulations in $M_1$ batches, each containing $M_2$ simulations, and use the notation $\mathbf{X}_{T,k,j}$ and $\mathbf{X}_{T,k,j}$ for the approximate and exact solution, respectively, given by the $k$th trajectory in the $j$th batch. For each batch we get average errors

$$\bar{\varepsilon}_j = \frac{1}{M_2} \sum_{k=1}^{M_2} |X_{T,k,j} - \mathbf{X}_{T,k,j}|,$$  \hfill (7.24)

which are independent and approximately Gaussian. Using the average errors $\bar{\varepsilon}_j$ we can estimate the mean

$$\bar{\varepsilon} = \frac{1}{M_1} \sum_{j=1}^{M_1} \bar{\varepsilon}_j = \frac{1}{M_1 M_2} \sum_{j=1}^{M_1} \sum_{k=1}^{M_2} |X_{T,k,j} - \mathbf{X}_{T,k,j}|,$$  \hfill (7.25)

and variance

$$\sigma^2_{\bar{\varepsilon}} = \frac{1}{M_1 - 1} \sum_{j=1}^{M_1} (\bar{\varepsilon}_j - \bar{\varepsilon})^2,$$  \hfill (7.26)

of the batch averages. For $M_2$ large enough, each batch is approximately Gaussian and (following the lines of Section 1.9 in [KP92]) a $100 (1 - \alpha) \%$ confidence interval for the absolute error is given by $[\bar{\varepsilon} - \Delta \bar{\varepsilon}, \bar{\varepsilon} + \Delta \bar{\varepsilon}]$ with

$$\Delta \bar{\varepsilon} = t_{1-\alpha,M_1-1} \sqrt{\frac{\sigma^2_{\bar{\varepsilon}}}{M_1}},$$  \hfill (7.27)

where $t_{1-\alpha,M_1-1}$ is determined from the Student $t$-distribution with $M_1 - 1$ degrees of freedom. It follows from (7.27) that the size of the confidence interval decreases as $M_1^{-\frac{1}{2}}$, which is quite slow.

To further investigate the origin of the error, we decompose the random variable $\hat{\varepsilon}$ into two parts

$$\hat{\varepsilon} = \varepsilon_{\text{disc}} + \varepsilon_{\text{stat}},$$  \hfill (7.28)

where $\varepsilon_{\text{disc}}$ is the time discretization error and $\varepsilon_{\text{stat}}$ is the statistical error. We define

$$\varepsilon_{\text{disc}} := E[\hat{\varepsilon}] = E \left[ \frac{1}{M} \sum_{k=1}^{M} |X_{T,k} - \mathbf{X}_{T,k}| \right] = E \left[ |X_T - \mathbf{X}_T| \right] = \varepsilon,$$  \hfill (7.29)

and hence the time discretization error coincides with the absolute error. This part of the error depends entirely on the choice of numerical method.
For a large number $M$ of independent simulations, the statistical error $\varepsilon_{\text{stat}} = \hat{\varepsilon} - \varepsilon_{\text{disc}}$ is, once more due to the Central Limit Theorem, approximately a Gaussian random variable with mean zero and variance

$$
\text{Var}(\varepsilon_{\text{stat}}) = E \left[ (\hat{\varepsilon} - \varepsilon_{\text{disc}})^2 \right] = E \left[ \left( \frac{1}{M} \sum_{k=1}^{M} |X_{T,k} - \mathbb{X}_{T,k}| - \bar{\varepsilon} \right)^2 \right]
$$

$$
= \frac{1}{M} E \left[ \left( \frac{1}{M} \sum_{k=1}^{M} (|X_{T,k} - \mathbb{X}_{T,k}| - \bar{\varepsilon}) \right)^2 \right] = \frac{1}{M^2} \sum_{k=1}^{M} E \left[ (|X_{T,k} - \mathbb{X}_{T,k}| - \bar{\varepsilon})^2 \right]
$$

$$
= \frac{1}{M} E \left[ (|X_T - \mathbb{X}_T| - \bar{\varepsilon})^2 \right] = \frac{1}{M} \text{Var} \left( |X_T - \mathbb{X}_T| \right),
$$

so the variance of the statistical error is inversely proportional to the total number $M$ of simulations. The choice of numerical method only affects the statistical error by means of the constant $\text{Var}(X_T)$.

In addition to the time discretization error and statistical error, there is always the risk that the result is influenced by roundoff errors, especially in the case of non-linear equations. From (7.30) it is clear, however, that a very large number of trajectories are required in order to make the statistical error small, so the roundoff error can often be neglected in comparison to the other two types of error. The effect of roundoff error is further discussed in Section 7.7 below.

From the discussion above we see that the choice of numerical method mainly affects the time discretization error. Hence, it is natural to use the convergence criterion (7.15) when comparing different numerical approximation methods. Note that the investigation of different types of errors found in this section can be carried out in exactly the same manner for weak approximations (see Section 9.4 in [KP92] for details).

For stochastic differential equations without explicit solutions, which constitute the ground for considering numerical methods, the discussion above is still valid but now $X_T$ is unknown, so the absolute error cannot be found numerically. To estimate the order of convergence we can proceed as follows. Generate a number of approximations $X_T^{(n)}$ using an increasing number $N^{(n)} = m \cdot 2^n$ of equidistant time steps. Under the (fairly strong) assumptions that numerical approximations with an increasing number of time steps approach the true solution monotonically and that the inequality in (7.15) can be replaced by an equality, we get the expansion

$$
E \left[ |X_T^{(n+1)} - X_T^{(n)}| \right] = E \left[ |X_T^{(n)} - X_T| \right] - E \left[ |X_T^{(n+1)} - X_T| \right]
$$

$$
= C \left( \frac{T}{m \cdot 2^n} \right) - C \left( \frac{T}{m \cdot 2^{n+1}} \right) = C \left( \frac{T}{m \cdot 2^n} \right)^\gamma \left( 1 - \frac{1}{2^\gamma} \right)
$$

so that approximately

$$
\frac{\log \frac{E \left[ |X_T^{(n+1)} - X_T^{(n)}| \right]}{N^{(n+2)}}}{\log \frac{N^{(n+2)}}{N^{(n+1)}}} = \frac{\log \frac{C \left( \frac{T}{m \cdot 2^n} \right)^\gamma \left( 1 - \frac{1}{2^\gamma} \right)}{C \left( \frac{T}{m \cdot 2^{n+1}} \right)^\gamma \left( 1 - \frac{1}{2^\gamma} \right)}}{\log \frac{m \cdot 2^{n+2}}{m \cdot 2^{n+1}}} = \frac{\log 2^\gamma}{\log 2} = \gamma.
$$

This estimate must of course be combined with a theoretical proof that the numerical solution actually converges to the correct solution.
7.4 Consistency and convergence

The concept of consistency can be used to show fairly general convergence results for time discrete approximations of stochastic differential equations. We say that the time discrete approximation $\overline{X}$ with maximum step size $\delta$ is strongly consistent if there exists a nonnegative function $c(\delta)$ such that

\begin{align}
(i) & \quad \lim_{\delta \to 0} c(\delta) = 0, \\
(ii) & \quad E \left[ \left( \frac{\overline{X}_{t_{n+1}} - \overline{X}_{t_n}}{\Delta t_n} \right) \right] - a(t_n, \overline{X}_{t_n}) \right)^2 \leq c(\delta), \\
(iii) & \quad E \left[ \left( \frac{1}{\Delta t_n} \right) \overline{X}_{t_{n+1}} - \overline{X}_{t_n} \right] - b(t_n, \overline{X}_{t_n}) \Delta W_{t_n} \right)^2 \leq c(\delta),
\end{align}

for all fixed values $\overline{X}_{t_n} = x$ and $n = 0, 1, ..., N - 1$. The second criterion in (7.33) asserts that the mean of the increment of the approximation converges to the mean of the increment of the stochastic differential $X$. The third criterion in (7.33) asserts that the variance of the difference between the random parts of the approximation and the random parts of the stochastic differential converges to zero. For autonomous stochastic differentials

\begin{equation}
dX_t = a(X_t) dt + b(X_t) dW_t,
\end{equation}

the following result holds.

**Theorem 7.2 (9.6.2 in [KP92])** Under the assumptions of Theorem 4.4, a strongly consistent equidistant time discrete approximation $\overline{X}$ of an autonomous stochastic differential $X$ with $\overline{X}_0 = X_0$ converges strongly to $X$.

**Proof.** See [KP92].

Similarly we say that the time discrete approximation $\overline{X}$ with maximum step size $\delta$ is weakly consistent if there exists a nonnegative function $c(\delta)$ such that

\begin{align}
(i) & \quad \lim_{\delta \to 0} c(\delta) = 0, \\
(ii) & \quad E \left[ \left( \frac{\overline{X}_{t_{n+1}} - \overline{X}_{t_n}}{\Delta t_n} \right) \right] - a(t_n, \overline{X}_{t_n}) \right)^2 \leq c(\delta), \\
(iii) & \quad E \left[ \left( \frac{1}{\Delta t_n} \right) \overline{X}_{t_{n+1}} - \overline{X}_{t_n} \right] - b(t_n, \overline{X}_{t_n}) \Delta W_{t_n} \right)^2 \leq c(\delta),
\end{align}

for all fixed values $\overline{X}_{t_n} = x$ and $n = 0, 1, ..., N - 1$. The first two criteria in (7.35) are the same as for strong consistency. The third criterion in (7.35), however, is much weaker than the third criterion in (7.33) and asserts only that the variance of the increments of the approximation should be close to the variance of the increments of the stochastic differential. For autonomous stochastic differentials the following result holds.

**Theorem 7.3 (9.7.4 in [KP92])** Suppose that the drift coefficient $a(x)$ and the diffusion coefficient $b(x)$ of (7.34) are four times continuously differentiable with polynomial growth and uniformly bounded derivatives. Let $\overline{X}$ be a weakly consistent equidistant time discrete approximation of the autonomous stochastic differential $X$ with $\overline{X}_0 = X_0$, such that

\begin{equation}
E \left[ \max_{0 \leq n \leq N} \left| \overline{X}_{t_n} \right|^{2q} \right] \leq K \left( 1 + E \left[ \left| X_0 \right|^{2q} \right] \right),
\end{equation}

68
for \( q = 1, 2, \ldots \), and
\[
E \left[ \frac{1}{\Delta t_n} |X_{t_{n+1}} - X_{t_n}|^6 \right] \leq c(\delta),
\] (7.37)
for \( n = 0, 1, \ldots, N - 1 \), where \( c(\delta) \) is the function in the definition of weak consistency. Then \( \overline{X} \) converges weakly to \( X \).

\textbf{Proof.} See [KP92]. \[ \Box \]

It is left as an exercise to show that the Euler approximation is strongly consistent, weakly consistent and satisfies conditions (7.36)-(7.37). As a consequence, the Euler approximation converges strongly and weakly.

### 7.5 Strong order of convergence for the Euler approximation

In Section 7.2 we mentioned that the Euler approximation converges strongly with order 0.5. Below we prove this result for most solvable stochastic differential equations.

**Theorem 7.4** Let \( X \) be a stochastic differential given by
\[
dX_t = a(t, X_t) \, dt + \sum_{k=1}^{m} b^{(k)}(t, X_t) \, dW_t^{(k)},
\] (7.38)
with initial data \( X_0 \) satisfying
\[
E \left[ |X_0|^2 \right] < \infty
\] (7.39)
\[
E \left[ |X_0 - \overline{X}_0|^2 \right] \leq K\delta,
\] (7.40)
where \( \overline{X}_0 \) is the initial data of the Euler approximation. Furthermore suppose that for some constant \( K \), independent of \( \delta \), the drift and diffusion coefficients satisfy, for all \( s, t \in [0, T] \), \( x, y \in \mathbb{R} \) and \( 1 \leq j \leq m \),

1. \( |a(t, x) - a(t, y)| + |b^{(j)}(t, x) - b^{(j)}(t, y)| \leq K |x - y| \), \quad (Lipschitz continuity)
2. \( |a(t, x)|^2 + |b^{(j)}(t, x)|^2 \leq K^2 \left( 1 + |x|^2 \right) \). \quad (Linear growth)
3. \( |a(s, x) - a(t, x)| + |b^{(j)}(s, x) - b^{(j)}(t, x)| \leq K |s - t|^{1/2} \). \quad (Hölder continuity of order \( \frac{1}{2} \))

Then there exists a constant \( C \), independent of \( \delta \), such that the equidistant Euler approximation \( \overline{X} \) satisfies
\[
E \left[ |X_T - \overline{X}_T| \right] \leq C\delta^{3/2},
\] (7.41)
for sufficiently small \( \delta \).

\textbf{Proof.} For theoretical purposes we define the so-called continuous Euler approximation \( \overline{X}(t) \) as
\[
\overline{X}(t) = \overline{X}_{t_{n-1}} + \int_{t_{n-1}}^{t} a(t_{n-1}, \overline{X}_{t_{n-1}}) \, dt + \sum_{k=1}^{m} \int_{t_{n-1}}^{t} b^{(k)}(t_{n-1}, \overline{X}_{t_{n-1}}) \, dW_t^{(k)},
\] (7.42)
for \( t \in [t_{n-1}, t_n] \), \( n = 0, 1, \ldots, N - 1 \). By construction, \( \overline{X} (t) \) will be continuous and \( \overline{X} (t_n) = \overline{X}_{t_n} \), for all \( n = 0, 1, \ldots, N - 1 \). Introduce \( Y_t = X_t - \overline{X} (t) \) and apply Itô’s formula to \( g (Y_t) = (Y_t)^2 \)

\[
d \left( (Y_t)^2 \right) = 2Y_t dY_t + \frac{1}{2} 2 (dY_t)^2 = 2 \left( X_t - \overline{X} (t) \right) (a (t, X_t) - a (t_{n-1}, \overline{X}_{t_{n-1}})) dt
\]

\[
+ 2 \left( X_t - \overline{X} (t) \right) \sum_{k=1}^{m} \left( b^{(k)} (t, X_t) - b^{(k)} (t_{n-1}, \overline{X}_{t_{n-1}}) \right) dW_t^{(k)}
\]

\[
+ \sum_{k=1}^{m} \left( b^{(k)} (t, X_t) - b^{(k)} (t_{n-1}, \overline{X}_{t_{n-1}}) \right)^2 dt.
\]

(7.43)

Now define

\[
Z (t) = E \left[ \left| X_t - \overline{X} (t) \right|^2 \vert A_0 \right] = E \left[ (Y_t)^2 \vert A_0 \right].
\]

(7.44)

Then (7.43) implies

\[
Z (t_n) = E \left[ (Y_{t_n})^2 \vert A_0 \right] = E \left[ (Y_{t_{n-1}})^2 \vert A_0 \right] + E \left[ \int_{t_{n-1}}^{t_n} d \left( (Y_s)^2 \right) \vert A_0 \right]
\]

\[
= E \left[ \left( X_{t_{n-1}} - \overline{X} (t_{n-1}) \right)^2 \vert A_0 \right]
\]

\[
+ E \left[ \int_{t_{n-1}}^{t_n} \left( X_s - \overline{X} (s) \right) (a (s, X_s) - a (t_{n-1}, \overline{X}_{t_{n-1}})) \, ds \vert A_0 \right]
\]

\[
+ \sum_{k=1}^{m} E \left[ \int_{t_{n-1}}^{t_n} \left( b^{(k)} (s, X_s) - b^{(k)} (t_{n-1}, \overline{X}_{t_{n-1}}) \right)^2 \, ds \vert A_0 \right]
\]

\[
\leq Z (t_{n-1}) + \int_{t_{n-1}}^{t_n} \left( E \left[ \left| X_s - \overline{X} (s) \right|^2 \vert A_0 \right] + E \left[ \left| a (s, X_s) - a (t_{n-1}, \overline{X}_{t_{n-1}}) \right|^2 \vert A_0 \right] \right) ds,
\]

(7.45)

where the last inequality follows by \( 2cd \leq c^2 + d^2 \). If we let \( f \) denote either \( a \) or \( b^{(j)} \) then by Lipschitz and Hölder continuity we have

\[
\left| f (s, X_s) - f (t_{n-1}, \overline{X}_{t_{n-1}}) \right|^2 \leq 3 \left| f (s, X_s) - f (s, X_{t_{n-1}}) \right|^2
\]

\[
+ 3 \left| f (s, X_{t_{n-1}}) - f (t_{n-1}, \overline{X}_{t_{n-1}}) \right|^2
\]

\[
+ 3 \left| f (t_{n-1}, \overline{X}_{t_{n-1}}) - f (t_{n-1}, \overline{X}_{t_{n-1}}) \right|^2
\]

\[
\leq 3K^2 \left| X_s - X_{t_{n-1}} \right|^2 + 3K^2 \left| s - t_{n-1} \right|
\]

\[
+ 3K^2 \left| X_{t_{n-1}} - \overline{X}_{t_{n-1}} \right|^2,
\]

(7.46)

so that

\[
Z (t_n) \leq Z (t_{n-1}) + \int_{t_{n-1}}^{t_n} \left( E \left[ \left| X_s - \overline{X} (s) \right|^2 \vert A_0 \right] \right) ds
\]

\[
+ C_1 \int_{t_{n-1}}^{t_n} \left( E \left[ \left| X_s - X_{t_{n-1}} \right|^2 \vert A_0 \right] + \left| s - t_{n-1} \right| \right) ds
\]

\[
+ C_1 \int_{t_{n-1}}^{t_n} E \left[ \left| X_{t_{n-1}} - \overline{X}_{t_{n-1}} \right|^2 \vert A_0 \right] \, ds.
\]

(7.47)
It is shown in Theorem 4.5.4 in [KP92] that under the assumptions in Theorem 4.4, there exists a constant $K_1$ depending only on $K$, $T$ and $X_0$ such that
\[
E \left[ |X_s - X_{t_{n-1}}|^2 \mid A_0 \right] \leq K_1 (s - t_{n-1}).
\] (7.48)

Hence (7.47) reduces to
\[
Z(t_n) \leq Z(t_{n-1}) (1 + C_1 \delta) + C_2 \int_{t_{n-1}}^{t_n} (s - t_{n-1}) \, ds + \int_{t_{n-1}}^{t_n} E \left[ (X_s - X(s))^2 \mid A_0 \right] \, ds
\]
\[
= Z(t_{n-1}) (1 + C_1 \delta) + C_3 \delta^2 + \int_{t_{n-1}}^{t_n} Z(s) \, ds.
\] (7.49)

An application of the Gronwall inequality with $\alpha(t) = Z(t)$, $\beta(t) = \beta = Z(t_{n-1}) (1 + C_1 \delta) + C_3 \delta^2$, $t_0 = t_{n-1}$, $T = t_n$ and $L = 1$ now yields, for $t = t_n$,
\[
Z(t_n) \leq \beta + \int_{t_{n-1}}^{t_n} \exp (t_n - s) \beta \, ds = \beta \left( 1 + [\exp (t_n - s)]_{t_{n-1}}^{t_n} \right) = \beta \exp (\delta)
\]
\[
= \left( Z(t_{n-1}) (1 + C_1 \delta) + C_3 \delta^2 \right) \exp (\delta) := A + BZ(t_{n-1}),
\] (7.50)
with $A = C_3 \delta^2 \exp (\delta)$ and $B = (1 + C_1 \delta) \exp (\delta)$. Iterating this upper bound, we obtain
\[
Z(t_n) \leq A + BZ(t_{n-1}) \leq A + B (A + BZ(t_{n-2})) \leq A + BA + B^2 (A + BZ(t_{n-3}))
\]
\[
\leq ... \leq A (1 + B + ... + B^{n-1}) + B^n Z(0) = A \frac{B^n - 1}{B - 1} + B^n Z(0).
\] (7.51)

Since the time discretization is equidistant, we have $n = \frac{T}{\delta}$, implying that
\[
\lim_{\delta \to 0^+} B^n = \lim_{\delta \to 0^+} \left( 1 + C_1 \delta \right) \frac{T}{\delta} \left( \exp (\delta) \right) \frac{T}{\delta} = \exp (T) \lim_{n=\frac{T}{\delta} \to \infty} \left( 1 + \frac{C_1 T}{n} \right)^n = \exp ((C_1 + 1) T).
\] (7.52)

Furthermore with the aid of l'Hôpital’s rule we get
\[
\lim_{\delta \to 0^+} \frac{A}{B - 1} = \lim_{\delta \to 0^+} \frac{C_3 \delta \exp (\delta)}{(1 + C_1 \delta) \exp (\delta) - 1} = \lim_{\delta \to 0^+} \frac{C_3 \exp (\delta) (\delta + 1)}{(C_1 \delta + C_1 \delta + 1)} = \frac{C_3}{1 + C_1}.
\] (7.53)

From (7.52)-(7.53) it is clear that there exists a $\delta_0$ such that, for all $\delta \in (0, \delta_0)$, one can find constants $C_4$, $C_5$ and $C_6$, independent of $\delta$, satisfying
\[
Z(t_n) \leq A \frac{1 - B^n}{1 - B} + B^n Z(0) \leq C_3 \delta + C_5 Z(0) \leq C_6 \delta,
\] (7.54)

where we used (7.40) in the last inequality. Jensen’s inequality (1.39) with $g(x) = x^2$ finally yields
\[
E \left[ |X_T - X_T| \right] \leq \sqrt{E \left[ (X_T - X(T))^2 \mid A_0 \right]} = \sqrt{E \left[ E \left[ (X_T - X(T))^2 \mid A_0 \right] \right]} \leq \sqrt{E[Z(T)]} \leq C_6 \delta^{\frac{1}{2}},
\] (7.55)
as desired. □

It is straightforward to generalize this result to multi-dimensional stochastic differentials $X$. Note also that the proof of the similar Theorem 10.2.2 in [KP92] asserts that the error bound in (7.41) holds uniformly on the interval $[0, T]$. This stronger result, however, requires the uniform mean square estimate for multiple Itô integrals proved in Lemma 10.8.1 in [KP92], which falls outside the scope of this course.
7.6 Weak order of convergence for the Euler approximation

In this section we show that the Euler approximation converges weakly with order 1, as was mentioned earlier in Section 7.2. We consider only the case of one-dimensional stochastic differentials and begin by defining the class of functions to which the drift and diffusion coefficients and function \( g \) should belong. Denote by \( \mathcal{H}_T^{(l)} \) the space of functions \( u : [0, T] \times \mathbb{R} \to \mathbb{R} \) such that the norm

\[
\|u\|_T^{(l)} = \sum_{2r+s \leq |l|} \sup_{t,x} \left| \frac{\partial^{r+s} u}{\partial t^r \partial x^s}(t, x) \right| + \sum_{2r+s = |l|} \sup_{t \neq t'} \frac{\partial^{r+s} u}{\partial t^r \partial x^s}(t, x) - \frac{\partial^{r+s} u}{\partial t^r \partial x^s}(t, x') \right| \frac{1}{|t-x|^{1-|l|}} + \sum_{0 < |2r-s| < 2 |t'-t|} \frac{\partial^{r+s} u}{\partial t^r \partial x^s}(t, x) - \frac{\partial^{r+s} u}{\partial t^r \partial x^s}(t', x) \right| ,
\]

(7.56)
is finite. Here \(|l|\) is the integer part of \( l \). \( \mathcal{H}_T^{(l)} \) contains functions whose partial derivatives \( \frac{\partial^{r+s} u}{\partial t^r \partial x^s} \) up to a certain order are Hölder continuous with index \( l - |l| \) in \( x \) and with index \( (l - 2r - s)/2 \) in \( t \). In the proof we use the following lemma.

Lemma 7.5 Let \( X \) be a stochastic differential given by

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

(7.57)

for bounded coefficients \( a \) and \( b \). Note that the noise is scalar. Let \( \overline{X}(t) \) be the continuous Euler approximation of \( X \) with maximal time step \( \delta \). Then there exists a positive constant \( K \), independent of \( \delta \), such that

\[
|E \left[ f(t, \overline{X}(t)) - f(t_{n(t)}, X_{n(t)}) \mid A_{n(t)} \right] | \leq K \|f\|_T^{(l)} \delta,
\]

(7.58)

for all \( t \in [0, T] \) and \( f \in \mathcal{H}_T^{(l)} \), with \( l \in (2, 3) \).

**Proof.** Since \( f \in \mathcal{H}_T^{(l)} \), with \( l \in (2, 3) \), the derivatives

\[
\frac{\partial f}{\partial t}, \frac{\partial f}{\partial x}, \text{ and } \frac{\partial^2 f}{\partial x^2},
\]

(7.59)
are uniformly bounded by \( L \|f\|_T^{(l)} \). Furthermore the derivatives in (7.59) are Hölder continuous and hence continuous, so we may apply Itô’s formula to \( f(t, \overline{X}(t)) \), arriving at

\[
df(t, \overline{X}(t)) = \frac{\partial f}{\partial t} \, dt + \frac{\partial f}{\partial x} \, d\overline{X}(t) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (d\overline{X}(t))^2.
\]

(6.60)

By the definition of \( \overline{X}(t) \) and the fact that the expectation of an Itô integral is zero, we obtain

\[
|E \left[ f(t, \overline{X}(t)) - f(t_{n(t)}, \overline{X}(t_{n(t)})) \mid A_{n(t)} \right] |
= E \left[ \int_{t_{n(t)}}^t \left( \frac{\partial f}{\partial t} + a(t_{n(t)}, \overline{X}(t_{n(t)})) \frac{\partial f}{\partial x} + \frac{1}{2} b(t_{n(t)}, \overline{X}(t_{n(t)}))^2 \frac{\partial^2 f}{\partial x^2} \right) \, d\overline{X}(t) \mid A_{n(t)} \right] 
\leq K \|f\|_T^{(l)} \delta,
\]

(7.61)
as desired. To prove the last inequality, we used the uniform bounds on the derivatives and the boundedness of \( a \) and \( b \).
Theorem 7.6 Let $X$ be a stochastic differential given by (7.57) with initial data $X_0$. Assume that all moments $E \|X_T\|^p$ of the solution exist and that

$$a, b \in \mathcal{H}_T^{l}, \quad g \in \mathcal{H}_T^{l+2}, \quad (7.62)$$

for some $l \in (2, 3)$. Then the Euler approximation $\tilde{X}$ with maximal time step $\delta$ satisfies

$$|E[g(X_T)] - E\left[ g\left(\tilde{X}_T\right)\right] | \leq C\delta, \quad (7.63)$$

for some positive constant $C$, independent of $\delta$.

Proof. From the Kolmogorov backward equation we know that

$$u(t, x) = E_t^T[g(X_T)], \quad (7.64)$$

solves the terminal value problem

$$\begin{cases}
\frac{\partial u}{\partial t} + Au = 0, & \text{for } t \in [0, T], \ x \in \mathbb{R} \\
u(T, x) = g(x), & \text{for } x \in \mathbb{R}
\end{cases} \quad (7.65)$$

with $A$ as in (5.18). Theorem 5.2 in [LSU68] furthermore states that the solution to this terminal value problem belongs to $\mathcal{H}_T^{l+2}$ and satisfies the regularity relation $\|u\|_{T}^{l+2} \leq L \|g\|_{T}^{l+2}$. From (7.64) we get

$$E\left[ g\left(\tilde{X}_T\right)\right] = E\left[ E_T^T X_T [g(X_T)]\right] = E\left[ u(T, X_T)\right], \quad (7.66)$$

$$E[g(X_T)] = E\left[ E_T^T X_T [g(X_T)]\right] = E\left[ u(0, X_0)\right], \quad (7.67)$$

allowing the discretization error to be represented as

$$|E[g(X_T)] - g(X_T)| = |E[u(T, X_T) - u(0, X_0)]| = \left| E\left[ \int_0^T du(t, \tilde{X}(t)) \right]\right|. \quad (7.68)$$

In analogy with (7.61), the differential $du(t, \tilde{X}(t))$ can be determined by Itô’s formula, taking the definition of $\tilde{X}(t)$ in (7.42) into account. Integrating and taking expectations, we obtain

$$E\left[ \int_0^T du(t, X(t)) \right] = E\left[ \int_0^T \left( \frac{\partial u}{\partial t} + a\left(t_n(t), X(t_n(t))\right) \frac{\partial u}{\partial x} + \frac{1}{2} \left( b\left(t_n(t), X(t_n(t))\right) \right)^2 \frac{\partial^2 u}{\partial x^2} \right) dt \right], \quad (7.69)$$

where we have suppressed the evaluation point $(t, \tilde{X}(t))$ in $u$ and its derivatives. Moreover replacing $\frac{\partial u}{\partial t}$ with $-Au$ according to (7.65), we arrive at

$$\frac{\partial u}{\partial t} + a\left(t_n(t), X(t_n(t))\right) \frac{\partial u}{\partial x} + \frac{1}{2} \left( b\left(t_n(t), X(t_n(t))\right) \right)^2 \frac{\partial^2 u}{\partial x^2} = \left( a\left(t_n(t), X(t_n(t))\right) - a(t, X(t))\right) \frac{\partial u}{\partial x} + \frac{1}{2} \left( b\left(t_n(t), X(t_n(t))\right) \right)^2 - \left( b(t, X(t)) \right)^2 \frac{\partial^2 u}{\partial x^2}. \quad (7.70)$$
so the expectation in (7.68) equals

\[
\left| E \left[ \int_0^T du \left( t, \overline{X} (t) \right) \right] \right| \\
\leq \left| E \left[ \int_0^T \left( a \left( t \left( t_n(t) \right) \right) - a \left( t, \overline{X} (t) \right) \right) \frac{\partial u}{\partial x} \left( t, \overline{X} (t) \right) dt \right] \right| \\
+ \left| E \left[ \int_0^T \frac{1}{2} \left( \left( b \left( t \left( t_n(t) \right) \right) \right)^2 - \left( b \left( t, \overline{X} (t) \right) \right)^2 \right) \frac{\partial^2 u}{\partial x^2} \left( t, \overline{X} (t) \right) dt \right] \right| \\
(7.71)
\]

From \( u \in H_T^{l+2} \) we conclude that \( \frac{\partial u}{\partial x} \in H_T^{l+1} \) and \( \frac{\partial^2 u}{\partial x^2} \in H_T^l \), so, due to the result in [LSU68], it is clear that the derivatives of \( u \) in (7.71) are bounded from above by \( L \| g \|^l_T \). Furthermore, since \( a, b \in H_T^l \) for some \( l \in (2, 3) \) (implying also that \( a \) and \( b \) are bounded) the expression

\[
\left| E \left[ \left( a \left( t \left( t_n(t) \right) \right) \right) - a \left( t, \overline{X} (t) \right) \right] \right| = \left| E \left[ \left( a \left( t \left( t_n(t) \right) \right) \right) - a \left( t, \overline{X} (t) \right) \right] \left| A_{n(t)} \right| \right| \\
(7.72)
\]

and the similar expression containing the diffusion coefficient satisfy the prerequisites of Lemma 7.5. Thus it immediately follows that

\[
E \left[ \left| g \left( \overline{X}_T \right) - g \left( X_T \right) \right| \right] = \left| E \left[ \int_0^T du \left( t, \overline{X} (t) \right) \right] \right| \leq TL \| g \|^l_T \left( K \| a \|^l_T \delta + K \| b \|^l_T \delta \right) \leq C\delta. \\
(7.73)
\]

This result is readily generalized to higher dimensions, but in that case the matrix \( b^{(l)}b^{(l)} \) must be uniformly elliptic (cf. Theorem 14.1.5 in [KP92]) to guarantee that the solution to the terminal value problem (7.65) has a solution with the stated regularity.

### 7.7 Stochastic numerical stability

Convergence of a numerical method is no guarantee that the method is effective for solving all kinds of stochastic differential equations. We must also ascertain that the propagation of initial errors and roundoff errors is kept under control. This is quantified by the concept of **numerical stability**. Let \( \overline{X}^{(1)} \) and \( \overline{X}^{(2)} \) be two numerical approximations of the solution to a stochastic differential equation corresponding to the initial values \( X_0^{(1)} \) and \( X_0^{(2)} \) respectively. Moreover, assume that \( \overline{X}^{(1)} \) and \( \overline{X}^{(2)} \) have been generated with the maximum step size \( \delta \). The numerical approximation is said to be **stochastically numerically stable** if for any \( \varepsilon > 0 \), there exists a \( \delta_0 > 0 \) such that for all step sizes \( \delta \in (0, \delta_0) \) it holds that

\[
\lim_{\left| X_0^{(1)} - X_0^{(2)} \right| \rightarrow 0} \sup_{0 \leq t \leq T} P \left( \left| \overline{X}^{(1)}_{t \left( t_n(t) \right)} - \overline{X}^{(2)}_{t \left( t_n(t) \right)} \right| \geq \varepsilon \right) = 0, \\
(7.74)
\]

for all stochastic differential equations for which the numerical approximation converges.

**Proposition 7.7** Let \( X \) be a stochastic differential satisfying the prerequisites of Theorem 4.4. Then the Euler approximation of \( X \) is stochastically numerically stable.
If we use this algorithm to approximate the geometric Brownian motion, we obtain

\[
\left| X_{t_{n(t)}}^{(1)} - X_{t_{n(t)}}^{(2)} \right|^2 \leq 3 \left| X_0^{(1)} - X_0^{(2)} \right|^2 + 3 \left( \int_0^{t_{n(t)}} \left( a\left( t_n(s), X_{t_n(s)}^{(1)} \right) - a\left( t_n(s), X_{t_n(s)}^{(2)} \right) \right) ds \right)^2 + 3 \left( \int_0^{t_{n(t)}} \left( b\left( t_n(s), X_{t_n(s)}^{(1)} \right) - b\left( t_n(s), X_{t_n(s)}^{(2)} \right) \right) dW_s \right)^2.
\]

Taking expectations of both sides, using Lipschitz continuity and the Itô isometry, we get

\[
E \left[ \left| X_{t_{n(t)}}^{(1)} - X_{t_{n(t)}}^{(2)} \right|^2 \right] \leq 3E \left[ \left| X_0^{(1)} - X_0^{(2)} \right|^2 \right] + 3K^2 T \int_0^{t_{n(t)}} E \left[ \left( X_{t_n(s)}^{(1)} - X_{t_n(s)}^{(2)} \right)^2 \right] ds
\]

\[
= 3E \left| X_0^{(1)} - X_0^{(2)} \right|^2 + L \int_0^{t_{n(t)}} E \left[ \left( X_{t_n(s)}^{(1)} - X_{t_n(s)}^{(2)} \right)^2 \right] ds.
\]

Taking supremum and applying the Gronwall inequality, we arrive at

\[
\sup_{0 \leq t \leq T} E \left[ \left| X_{t_{n(t)}}^{(1)} - X_{t_{n(t)}}^{(2)} \right|^2 \right] \leq \left( 3 + L \int_0^T \exp (L (T - s)) ds \right) E \left[ \left| X_0^{(1)} - X_0^{(2)} \right|^2 \right].
\]

Chebyshev's inequality \((1.41)\) finally implies

\[
P\left( \left| X_{t_{n(t)}}^{(1)} - X_{t_{n(t)}}^{(2)} \right| \geq \epsilon \right) = P\left( \left| X_{t_{n(t)}}^{(1)} - X_{t_{n(t)}}^{(2)} \right|^2 \geq \epsilon^2 \right) \leq \frac{1}{\epsilon^2} E \left[ \left| X_{t_{n(t)}}^{(1)} - X_{t_{n(t)}}^{(2)} \right|^2 \right],
\]

so

\[
\lim_{\left| X_0^{(1)} - X_0^{(2)} \right| \to 0} \sup_{0 \leq t \leq T} P\left( \left| X_{t_{n(t)}}^{(1)} - X_{t_{n(t)}}^{(2)} \right| \geq \epsilon \right) \leq \lim_{\left| X_0^{(1)} - X_0^{(2)} \right| \to 0} \sup_{0 \leq t \leq T} \frac{1}{\epsilon^2} E \left[ \left| X_{t_{n(t)}}^{(1)} - X_{t_{n(t)}}^{(2)} \right|^2 \right]
\]

\[
\leq \lim_{\left| X_0^{(1)} - X_0^{(2)} \right| \to 0} \frac{K}{\epsilon^2} E \left[ \left| X_0^{(1)} - X_0^{(2)} \right|^2 \right] = 0.
\]

and the stochastic numerical stability follows. \(\blacksquare\)

As an example of a time discrete approximation that is not stochastically numerically stable, we consider the fully implicit Euler approximation

\[
X_{t_{n+1}} = X_{t_n} + a\left( t_{n+1}, X_{t_{n+1}} \right) (t_{n+1} - t_n) + b\left( t_{n+1}, X_{t_{n+1}} \right) (W_{t_{n+1}} - W_{t_n}).
\]

If we use this algorithm to approximate the geometric Brownian motion, we obtain

\[
X_{t_{n+1}} = X_{t_n} + aX_{t_{n+1}} (t_{n+1} - t_n) + bX_{t_{n+1}} (W_{t_{n+1}} - W_{t_n}),
\]

from which \(X_{t_{n+1}}\) can be extracted as

\[
X_{t_{n+1}} = X_{t_n} \frac{1}{1 - a (t_{n+1} - t_n) - b (W_{t_{n+1}} - W_{t_n})}
\]

\[
= X_{t_0} \prod_{k=0}^{n} \frac{1}{1 - a (t_{k+1} - t_k) - b (W_{t_{k+1}} - W_{t_k})}.
\]
Apart from the case with very strong drift and weak noise intensity, this approximation is unsuitable, since one or several of the terms in the product may be infinite.

Note that in the case of weak approximations where only the closeness of distribution matters, the unbounded Wiener process may be replaced with a bounded process with similar statistical properties. Hence the fully implicit Euler approximation can, regardless of the fact that is not stochastically numerically stable, still be useful for weak approximations.

8 Itô–Taylor expansion

In this section we derive a stochastic analogue to the Taylor formula which in theory makes it possible to construct numerical approximations with arbitrarily high order of convergence. Consider the stochastic differential

$$X_t = X_0 + \int_0^t a(s, X_s) \, ds + \sum_{k=1}^m \int_0^t b^{(k)}(s, X_s) \, dW_s^{(k)}, \quad (8.1)$$

for $t \in [0, T]$, where the coefficients $a$ and $b$ are sufficiently smooth, real-valued functions satisfying a linear growth bound. We define the differential operators $L^0$ and $L^k$ as

$$L^0 = \frac{\partial}{\partial t} + a \frac{\partial}{\partial x} + \frac{1}{2} \left( \sum_{k=1}^m \left( b^{(k)}(s, X_s) \right)^2 \right) \frac{\partial^2}{\partial x^2}, \quad (8.2)$$

$$L^k = b^{(k)} \frac{\partial f}{\partial x}. \quad (8.3)$$

Then by the Itô formula we have, for any twice continuously differentiable function $f : [0, T] \times \mathbb{R} \to \mathbb{R}$,

$$f(t, X_t) = f(0, X_0) + \int_0^t \left( \frac{\partial f}{\partial t}(s, X_s) + a(s, X_s) \frac{\partial f}{\partial x}(s, X_s) + \frac{1}{2} \left( \sum_{k=1}^m \left( b^{(k)}(s, X_s) \right)^2 \right) \frac{\partial^2 f}{\partial x^2}(s, X_s) \right) \, dt + \sum_{k=1}^m \int_0^t b^{(k)}(s, X_s) \frac{\partial f}{\partial x}(s, X_s) \, dW_s^{(k)}$$

$$= f(0, X_0) + \int_0^t L^0 f(s, X_s) \, ds + \sum_{k=1}^m \int_0^t L^k f(s, X_s) \, dW_s^{(k)}. \quad (8.4)$$

Using (8.4) with $f(t, X_t) = a(t, X_t)$ and $f(t, X_t) = b(t, X_t)$, respectively, on (8.1), we obtain the following expansion for $X_t$

$$X_t = X_0 + \int_0^t \left( a(0, X_0) + \int_0^s L^0 a(u, X_u) \, du + \sum_{l=1}^m \int_0^s L^l a(u, X_u) \, dW_u^{(l)} \right) \, ds$$

$$+ \sum_{k=1}^m \int_0^t \left( b^{(k)}(0, X_0) + \int_0^s L^0 b^{(k)}(u, X_u) \, du + \sum_{l=1}^m \int_0^s L^l b^{(k)}(u, X_u) \, dW_u^{(l)} \right) dW_s^{(k)}$$

$$= X_0 + a(0, X_0) \int_0^t ds + \sum_{k=1}^m b^{(k)}(0, X_0) \int_0^t dW_s^{(k)} + R, \quad (8.5)$$
where the remainder $R$,

$$R = \int_0^t \int_0^s L^0 a (u, X_u) \, du \, ds + \sum_{l=1}^m \int_0^t \int_0^s L^l a (u, X_u) \, dW_u^{(l)} \, ds$$

$$+ \sum_{k=1}^m \int_0^t \int_0^s L^0 b^{(k)} (u, X_u) \, du \, dW_s^{(k)} + \sum_{k=1}^m \sum_{l=1}^m \int_0^t \int_0^s L^l b^{(k)} (u, X_u) \, dW_u^{(l)} \, dW_s^{(k)}$$

contains multiple stochastic integrals of order two. Note that without the remainder term, (8.5) reduces to the Euler approximation of $X$.

This expansion procedure can now be iterated. If we expand $f = L^1 b^{(k)}$ in the rest term of (8.5), we obtain

$$X_t = X_0 + a (0, X_0) \int_0^t ds + \sum_{k=1}^m b^{(k)} (0, X_0) \int_0^t dW_s^{(k)}$$

$$+ \sum_{k=1}^m \sum_{l=1}^m L^l b^{(k)} (0, X_0) \int_0^t \int_0^s dW_u^{(l)} \, dW_s^{(k)} + R',$$

with another rest term $R'$ containing multiple stochastic integrals of orders two and three. Removing $R'$, we get the so-called Milstein approximation, discussed in Section 9.1, which has strong order of convergence equal to 1.

Note that expansions such as (8.5) and (8.7) can easily be generalized to vector valued stochastic integrals if we, according to the multi-dimensional Ito formula, define the operators $L^0$ and $L^k$ as in (5.3.1)-(5.3.2) of [KP92].

These expansions will only be useful if we can find an interpretation to multiple stochastic integrals. For this we need a couple of definitions. We call a row vector

$$\alpha = (k_1, k_2, ..., k_l),$$

where $k_i \in \{0, 1, ..., m\}$ for $i \in \{1, 2, ..., l\}$ and $m = 1, 2, 3, ..., a multi-index$ of length $l := l (\alpha) \in \{1, 2, ...\}$. For completeness we let $v$ denote the multi-index with zero length. In addition we shall write $n (\alpha)$ for the number of zero components of the multi-index $\alpha$. Denote the set of all multi-indices by $\mathcal{M}$. Given $\alpha \in \mathcal{M}$ with $l (\alpha) \geq 1$, we write $-\alpha$ and $\alpha-$ for the multi-indices in $\mathcal{M}$ obtained by deleting the first and last components, respectively, of $\alpha$.

Let $\rho$ and $\tau$ be two stopping times such that $0 \leq \rho (t) \leq \tau (t) \leq T$ w.p.1. Then for a multi-index $\alpha = (k_1, k_2, ..., k_l) \in \mathcal{M}$ and a process $f \in \mathcal{H}_\alpha$ (where $\mathcal{H}_\alpha$ will be defined shortly), we define the multiple Ito integral $I_\alpha [f (\cdot)]_{\rho, \tau}$ recursively by

$$I_\alpha [f (\cdot)]_{\rho, \tau} := \begin{cases} f (\tau), & \text{for } l = 0, \\ \int_\rho^\tau I_{\alpha-} [f (\cdot)]_{\rho, s} \, ds, & \text{for } l \geq 1 \text{ and } k_l = 0, \\ \int_\rho^\tau I_{\alpha-} [f (\cdot)]_{\rho, s} \, dW_s^{(k_l)}, & \text{for } l \geq 1 \text{ and } k_l \geq 1. \end{cases}$$

For $\alpha = (k_1, k_2, ..., k_l)$ with $l \geq 2$, $\mathcal{H}_\alpha$ is the space of adapted right continuous processes with left hand limits such that the integral process $I_{\alpha-} [f (\cdot)]_{\rho, t}$ considered as a function of $t$ satisfies

$$I_{\alpha-} [f (\cdot)]_{\rho, t} \in \mathcal{H}_{(k_l)}.$$  (8.10)
Here $\mathcal{H}_0$ contains all processes such that $|f\left(t, \omega\right)| < \infty$ w.p.1, $\mathcal{H}_d$ contains all processes such that $\int_0^t |f\left(s, \omega\right)|\ ds < \infty$ and $\mathcal{H}_e$ contains all processes such that $\int_0^t |f\left(s, \omega\right)|^2\ ds < \infty$. As an example, we give two examples of multiple Itô integrals

$$I^{(1)}_t \left[f\left(\cdot\right)\right]_{0,t} = \int_0^t f\left(s_1\right)\ dW^{(1)}_{s_1},$$

$$I_{(0,1,2)}^{(1)} \left[f\left(\cdot\right)\right]_{0,t} = \int_0^t \int_0^s \int_0^{s_3} f\left(s_1\right)\ ds_1\ dW^{(1)}_{s_2}\ dW^{(2)}_{s_3}. \quad (8.12)$$

For multiple Itô integrals $I_{\alpha} \left[f\left(\cdot\right)\right]_{0,T}$ with $f \equiv 1$, which we for simplicity denote by $I_{\alpha}$, one can use Itô’s formula (for a proof see [KP92]) to derive the following relation.

**Proposition 8.1 (5.2.3 in [KP92])** Let $\alpha = (k_1, ..., k_l) \in \mathcal{M}$ with $k_1, ..., k_l \in \{0, 1, ..., m\}$ and $m = 1, 2, 3, ...$ Then

$$W^{(k)}_t I_{\alpha} = \sum_{i=0}^l I_{(k_1, ..., k_i, k_{i+1}, ..., k_l)} + \sum_{i=1}^l \chi_{\{k_i = k \neq 0\}} I_{(k_1, ..., k_{i-1}, 0, k_{i+1}, ..., k_l)}. \quad (8.13)$$

To illustrate the use of Proposition 8.1, we give two examples of formulas that can be deduced from (8.13) in the case of $m = 1$

$$W_t = W^{(1)}_t I_{(1,0)} = I_{(1,0)} + I_{(0,1)}, \quad (8.14)$$

$$W_t \int_0^t s\ dW_s = W^{(1)}_t I_{(0,1)} = I_{(1,0,1)} + I_{(0,1,1)} + I_{(0,0,1)} = 2I_{(0,1,1)} + I_{(1,0,1)} + I_{(0,0)}. \quad (8.15)$$

Both these equations will come in handy in the next chapter. For multi-indices $\alpha$ containing only one type of components, it is shown in [KP92] that Proposition 8.1 reduces to the recursive relation

**Corollary 8.2 (5.2.4 in [KP92])** Let $\alpha = (k_1, ..., k_l) \in \mathcal{M}$ with $l \geq 2$ and $k_1 = ... = k_l = k \in \{0, 1, ..., m\}$. Then

$$I_{\alpha} = \begin{cases} \frac{l!}{n!} I_n^{(k)} I_{\alpha-} - tI_{(\alpha-)} & \text{for } k = 0, \\ \frac{l!}{n!} I_n^{(k)} I_{\alpha-} - tI_{(\alpha-)} & \text{for } k \geq 1. \end{cases} \quad (8.16)$$

where the case $k \geq 1$ can be rewritten in terms of Hermite polynomials as

$$I_{(k,k)} = \frac{1}{2t} \left(I^{2}_{(k)} - t\right), \quad (8.17)$$

$$I_{(k,k,k)} = \frac{1}{3t} \left(I^{3}_{(k)} - 3tI^{2}_{(k)}\right), \quad (8.18)$$

$$I_{(k,k,k,k)} = \frac{1}{4t} \left(I^{4}_{(k)} - 6tI^{3}_{(k)} + 3t^{2}\right). \quad (8.19)$$

and so on.

From (8.7) it follows that in addition to multiple Itô integrals, the stochastic Taylor expansion will contain functions evaluated at the origin. Guided by this observation, we recursively define,
for each $\alpha = (k_1, \ldots, k_l)$ and function $f \in C^h\left(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R}\right)$ with $h = l(\alpha) + n(\alpha)$, the Itô coefficient function by

$$f_\alpha = \begin{cases} f, & \text{for } l = 0, \\ L^k f_{-\alpha}, & \text{for } l \geq 1. \end{cases}$$  \hspace{1cm} (8.20)$$

In the autonomous 1-dimensional case $d = m = 1$ we get the following Itô coefficient functions for $f(t, x) = x$

$$f(0) = L^0 f_v = L^0 f = a,$$
$$f(1) = L^1 f_v = L^1 f = b,$$
$$f(0,1) = L^0 f(1) = L^0 b = ab + \frac{1}{2} b^2 b'',$$
$$f(1,1) = L^1 f(1) = L^1 b = bb',$$

and so on, where the prime represents differentiation with respect to the $x$ variable.

The multiple stochastic integrals appearing in stochastic Taylor expansions cannot be chosen arbitrarily. On the contrary they must be chosen from a so-called hierarchical set. A subset $\mathcal{A} \subset \mathcal{M}$ is hierarchical if $\mathcal{A}$ is non-empty, if the lengths of the multi-indices in $\mathcal{A}$ are uniformly bounded, that is if

$$\sup_{\alpha \in \mathcal{A}} l(\alpha) < \infty,$$  \hspace{1cm} (8.25)$$

and if

$$-\alpha \in \mathcal{A} \text{ for each } \alpha \in \mathcal{A} \setminus \{v\}.$$  \hspace{1cm} (8.26)$$

Hence if a multi-index $\alpha$ belongs to a hierarchical set, then so does the multi-index $-\alpha$. As an example the sets

$$\{v\}, \{v, (0)\}, \{v, (1)\}, \{v, (0), (1)\}, \{v, (0), (1), (1, 1)\}$$

are all hierarchical.

If we form a stochastic Taylor expansion with multiple Itô integrals with multi-indices taken from a given hierarchical set $\mathcal{A}$, the remainder term will contain multiple Itô integrals with multi-indices taken from the remainder set $\mathcal{B}(\mathcal{A})$ corresponding to $\mathcal{A}$. The remainder set contains all the following multi-indices with respect to the given hierarchical set, and is hence given by

$$\mathcal{B}(\mathcal{A}) = \{\alpha \in \mathcal{M} \setminus \mathcal{A} : -\alpha \in \mathcal{A}\}.$$  \hspace{1cm} (8.28)$$

For example, for $m = 1$ we have

$$\mathcal{B}(\{v, (0), (1)\}) = \{(0, 0), (1, 0), (0, 1), (1, 1)\},$$

which corresponds to the expansion in (8.5)-(8.6).

We are now ready to state the Itô-Taylor expansion.

**Theorem 8.3 (Itô-Taylor expansion, 5.5.1 in [KP92])** Let $X_t$ be a $d$-dimensional stochastic differential

$$X_t^{(i)} = X_{t_0}^{(i)} + \int_{t_0}^t a^{(i)}(s, X_s) \, ds + \sum_{k=1}^{m} \int_{t_0}^t b^{(i,k)}(s, X_s) \, dW^{(k)}_s,$$  \hspace{1cm} (8.30)$$

and so on, where the prime represents differentiation with respect to the $x$ variable.
with $1 \leq i \leq d$ and $t \in [t_0, T]$. Let $\rho$ and $\tau$ be two stopping times with
\[ t_0 \leq \rho(\omega) \leq \tau(\omega) \leq T, \tag{8.31} \]
w.p.1, let $\mathcal{A} \subset \mathcal{M}$ be an hierarchical set and let $f: \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}$ be some function. Then the Itô-Taylor expansion
\[ f(\tau, X_\tau) = \sum_{\alpha \in \mathcal{A}} I_\alpha [f_\alpha (\rho, X_\rho)]_{\rho, \tau} + \sum_{\alpha \in \mathcal{B}(\mathcal{A})} I_\alpha [f_\alpha (\cdot, X)]_{\rho, \tau}, \tag{8.32} \]
holds provided that all derivatives of $f, a$ and $b$ and all multiple Itô integrals appearing in (8.32) exist.

The expansion is proved by repeated applications of Itô’s formula for stopping times and is not difficult, but has been left out (see pages 184-187 in [KP92]). In the following chapters we shall use Itô-Taylor expansions with suitable hierarchical sets to construct strong and weak numerical approximations to stochastic differential equations. Before pursuing that route, we demonstrate that under appropriate choices of the hierarchical set $\mathcal{A}$, the Itô-Taylor expansion reduces to the Itô formula or the Taylor formula, respectively.

To find the Taylor formula consider the case of $d = m = 1$, $f(t, x) = f(x) \in C^\infty (\mathbb{R}, \mathbb{R})$, $a \equiv 1$, $b \equiv 0$, $\rho = 0$ and $\tau = t \in [0, T]$, so that the stochastic differential in (8.30) reduces to $X_t = t$. With this choice the Itô coefficient functions vanish for all multi-indices containing one or more non-zeros components and by (8.2),(8.3) and (8.20)
\[ f(0) + \int_0^t 1 \frac{\partial f}{\partial x^i} (s, X_\omega) \, ds \leq \int_0^t \, ds, \tag{8.33} \]
For each $l = 0, 1, 2, \ldots$ we take the hierarchical set
\[ \Gamma_l = \{ \alpha \in \mathcal{M} : l(\alpha) \leq l \}, \tag{8.34} \]
with the corresponding remainder set $\mathcal{B}(\Gamma_l) = \{ \alpha \in \mathcal{M} : l(\alpha) = l + 1 \}$. Since the Itô coefficient functions vanish for non-zero components in the multi-indices, the only multi-indices in $\Gamma_l$ that will remain in the Itô-Taylor expansion are
\[ \Gamma_l = \{ \alpha = (0, \ldots, 0) : l(\alpha) \leq l \}, \tag{8.35} \]
with $\mathcal{B}(\Gamma_l) = \{ \alpha = (0, \ldots, 0) : l(\alpha) = l + 1 \}$. Then the Itô-Taylor expansion reduces to
\[ f(t) = f(X_t) = f(\tau, X_\tau) = \sum_{\alpha \in \Gamma_l} I_\alpha [f_\alpha (\rho, X_\rho)]_{\rho, \tau} + \sum_{\alpha \in \mathcal{B}(\Gamma_l)} I_\alpha [f_\alpha (\cdot, X)]_{\rho, \tau} \]
\[ = f(\rho, X_\rho) + \sum_{i=1}^{l} \int_0^\tau \int_0^{s_i} \cdots \int_0^{s_2} \frac{\partial f}{\partial x^i} (\rho, X_\rho) \, ds_1 \cdots ds_i \]
\[ + \sum_{i=1}^{l} \int_0^\tau \int_0^{s_{l+i-1}} \cdots \int_0^{s_2} \frac{\partial^{l+1} f}{\partial x^{l+1}} (\cdot, X_\omega) \, ds_1 \cdots ds_l \, ds_{l+1} \]
\[ = f(0) + \sum_{i=1}^{l} \frac{1}{l!} \frac{\partial f}{\partial x^i} (0) t^i + \int_0^t \int_0^{s_{l+i-1}} \cdots \int_0^{s_2} \frac{\partial^{l+1} f}{\partial x^{l+1}} (s_1) \, ds_1 \cdots ds_{l+1}, \tag{8.36} \]
where we have used (4.68) repeatedly in the last step. This is Taylor’s formula with remainder in integral form.

To find the Itō formula, consider the case of \( m \)-dimensional noise and choose \( A = \{ v \} \) implying \( B(A) = \{(0), (1), \ldots, (m)\} \). Then (8.32) takes the form

\[
f(\tau, X_\tau) = I_v [f_v (\rho, X_\rho)]_{\rho, \tau} + \sum_{\alpha \in B(v)} I_\alpha [f_\alpha (\cdot, X_\cdot)]_{\rho, \tau} + \int_0^\tau L^0 f (s, X_s) \, ds + \sum_{k=1}^m \int_0^\tau L^k f (s, X_s) \, dW^k_s.
\]

Now applying (8.2)-(8.3) to (8.37), we get the multidimensional Itō formula (3.17) with \( d = 1 \), so the Itō formula is an Itō-Taylor expansion, in which the remainder only contains multiple Itō integrals of multiplicity one.

## 9 Strong approximation

In this chapter we study various strong approximations of stochastic differential equations. The approximations considered belong to one of the following three categories: approximations based on Itō-Taylor expansions, approximations based on derivative-free versions of Itō-Taylor expansions and implicit approximations. In this section we shall only consider approximations of stochastic differentials with \( d = m = 1 \).

### 9.1 Milstein approximation

Apart from the Euler approximation, which we have already studied in detail, the simplest approximation corresponding to an Itō-Taylor expansion is the Milstein approximation already mentioned in connection to (8.7). It is given by

\[
\overline{X}_{t_{n+1}} = \overline{X}_t + a (t_n, \overline{X}_{t_n}) \Delta t_n + b (t_n, \overline{X}_{t_n}) \Delta W_{t_n} + b (t_n, \overline{X}_{t_n}) \frac{\partial b}{\partial x} (t_n, \overline{X}_{t_n}) \frac{(\Delta W_{t_n})^2 - \Delta t_n}{2}.
\]

Observe that by (8.17) \( I_{(1,1)} = \frac{1}{2} \left( (\Delta W_{t_n})^2 - t \right) \) and by (8.24) \( f_{(1,1)} = b b' \), so the last term in (9.1) equals \( f_{(1,1)} I_{(1,1)} \) and consequently the Milstein approximation contains the non-remainder terms of the Itō-Taylor expansion with hierarchical set \( A = \{ v, (0), (1), (1,1) \} \).

It follows from the theorem on strong order of convergence for approximations based on Itō-Taylor expansions stated in the next section that the strong order of convergence of the Milstein approximation is 1, so in a sense the Milstein approximation is the proper stochastic generalization of the Euler approximation for ordinary differential equation which has order of convergence equal to 1 as well.

### 9.2 Higher-order strong Itō-Taylor approximation

Theorem 10.6.3 of [KP92] states that under some regularity conditions on the Itō coefficient functions such as Lipschitz continuity, linear growth and continuity of the derivatives up to...
some given order, the approximation obtained by choosing the hierarchical set in the Itô-Taylor expansion as

$$\mathcal{A}_\gamma = \left\{ \alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq 2\gamma \text{ or } l(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \right\}, \quad (9.2)$$

for $\gamma = 0.5, 1, 1.5, \ldots$ converges strongly with order $\gamma$.

Determining $\mathcal{A}_\gamma$ for the lowest choices of $\gamma$, we obtain

$$\mathcal{A}_{0.5} = \left\{ \alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq 1 \text{ or } l(\alpha) = n(\alpha) = 1 \right\} = \{v, (0), (1)\}, \quad (9.3)$$

$$\mathcal{A}_1 = \left\{ \alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq 2 \text{ or } l(\alpha) = n(\alpha) = 1.5 \right\} = \{v, (0), (1), (1, 1)\}, \quad (9.4)$$

$$\mathcal{A}_{1.5} = \left\{ \alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq 3 \text{ or } l(\alpha) = n(\alpha) = 2 \right\}
= \{v, (0), (1), (0, 0), (0, 1), (1, 0), (1, 1), (1, 1, 1)\}, \quad (9.5)$$

where the first two hierarchical sets agrees with the Euler and Milstein approximations respectively. The last set corresponds to the order 1.5 strong Itô-Taylor scheme, due to Platen and Wagner, which in the autonomous case is given by (the first Itô coefficient functions for $f(t, x) = x$ can be found in (8.21)-(8.24))

$$\tilde{X}_{tn+1} = \sum_{\alpha \in \mathcal{A}_{1.5}} I_{\alpha} \left[ (\tilde{X}_{tn})_\alpha \right]_{tn, tn+1} + \sum_{\alpha \in \mathcal{A}_{1.5}} (\tilde{X}_{tn})_\alpha I_{\alpha}
= \tilde{X}_{tn} + aI_{(0)} + bI_{(1)} + \left( aa' + \frac{1}{2} b^2 a'' \right) I_{(0,0)} + \left( ab' + \frac{1}{2} b'^2 b'' \right) I_{(0,1)} + ba' I_{(1,0)}
+ bb'I_{(1,1)} + b \left( (b')^2 + bb'' \right) I_{(1,1,1)}, \quad (9.6)$$

where $a$, $b$ and the derivatives of these functions are evaluated at $\tilde{X}_{tn}$.

For this scheme to be of any use we must be able to calculate the multiple Itô integrals in (9.6) or at least find a way to simulate them. From the Milstein approximation we know that

$$I_{(0)} = \Delta t_n, \quad I_{(1)} = \Delta W_{tn}, \quad I_{(1,1)} = \frac{(\Delta W_{tn})^2 - \Delta t_n}{2}, \quad (9.7)$$

and it is also clear that

$$I_{(0,0)} = \int_{tn}^{tn+1} du ds = \frac{(\Delta t_n)^2}{2} \quad \text{and} \quad I_{(1,1,1)} = \frac{\Delta W_{tn}}{2} \left( \frac{(\Delta W_{tn})^2}{3} - \Delta t_n \right), \quad (9.8)$$

where the first formula follows from (8.16) and the second from (8.18). It remains to consider $I_{(0,1)}$ and $I_{(1,0)}$. We introduce the random variable $\Delta Z_{tn} := I_{(1,0)}$. Then it follows from (8.14) that $I_{(0,1)}$ can be expressed as

$$I_{(0,1)} = I_{(0)} I_{(1)} - I_{(1,0)} = \Delta t_n \Delta W_{tn} - \Delta Z_{tn}, \quad (9.9)$$

so if we can calculate $\Delta Z_{tn}$, we would be able to express both the remaining multiple Itô integrals. Unfortunately all multiple Itô integrals cannot be calculated explicitly. Nevertheless, due to Proposition 8.1, Corollary 8.2, the Itô isometry and the fact that the expectation of Itô integrals is zero, it is possible to determine the expectation and covariance of any multiple Itô integral. $\Delta Z_{tn}$ is a Gaussian variable and, hence, we can simulate the pair $(\Delta W_{tn}, \Delta Z_{tn})$ as two
joint Gaussian variables with the correct expectation and covariance. Observe that since the expectation of an Itô integral with adapted integrand is zero, the expectation of any multiple Itô integral with multi-index \( \alpha = (k_1, ..., k_{l-1}, k_l) \), with \( k_l \neq 0 \), is zero. Hence we get

\[
E[\Delta Z_{t_n}] = E[I_{(1,0)}] = E[\Delta t_n I_{(1)} - I_{(0,1)}] = 0,
\]

where we have used (8.14) to reformulate \( I_{(1,0)} \) is terms of functions whose expectation we can determine. Similarly

\[
E[(\Delta Z_{t_n})^2] = E[I_{(1,0)}^2] = E[(\Delta t_n I_{(1)} - I_{(0,1)})^2] = (\Delta t_n)^2 E[I_{(1)}^2] - 2\Delta t_n E[I_{(1)}] E[I_{(0,1)}] + E[I_{(0,1)}^2].
\]

(9.11)

We determine the three terms of (9.11) separately, applying (8.15) and (9.8) to the mid-term

\[
E[I_{(1)}^2] = E[(\Delta W_{t_n})^2] = \Delta t_n,
\]

(9.12)

\[
E[I_{(1)} I_{(0,1)}] = E[2I_{(0,1,1)} + I_{(1,0,1)} + I_{(0,0)}] = E[I_{(0,0)}] = \frac{(\Delta t_n)^2}{2},
\]

(9.13)

\[
E[I_{(0,1)}^2] = E\left[ \left( \int_{t_n}^{t_{n+1}} \int_{t_n}^{s} dr dW_s \right)^2 \right] = \int_{t_n}^{t_{n+1}} (s - t_n)^2 ds = \frac{(\Delta t_n)^3}{3}.
\]

(9.14)

so combining (9.11)-(9.14), we obtain

\[
E[(\Delta Z_{t_n})^2] = (\Delta t_n)^2 \Delta t_n - 2\Delta t_n \frac{(\Delta t_n)^2}{2} + \frac{(\Delta t_n)^3}{3} = \frac{3}{3}.
\]

(9.15)

Finally

\[
E[\Delta W_{t_n} \Delta Z_{t_n}] = E[I_{(1)} (\Delta t_n I_{(1)} - I_{(0,1)})] = \Delta t_n E[I_{(1)}^2] - E[I_{(1)} I_{(0,1)}]
\]

\[
= (\Delta t_n)^2 - \frac{(\Delta t_n)^2}{2} = \frac{(\Delta t_n)^2}{2}.
\]

(9.16)

Hence \((\Delta W_{t_n}, \Delta Z_{t_n})\) can be simulated from two independent \( N(0, 1) \) variables \( U_1 \) and \( U_2 \) by means of the transformation

\[
\Delta \tilde{W}_{t_n} = \sqrt{\Delta t_n} U_1 \quad \Delta \tilde{Z}_{t_n} = \frac{1}{2} (\Delta t_n)^2 \left( U_1 + \frac{1}{\sqrt{3}} U_2 \right).
\]

(9.17)

It is easy to see that this choice of \((\Delta \tilde{W}_{t_n}, \Delta \tilde{Z}_{t_n})\) has the correct expectation and covariance. Conclusively the order 1.5 strong Itô-Taylor scheme is given by

\[
\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + a\Delta t_n + b\Delta \tilde{W}_{t_n} + \left( a a' + \frac{1}{2} b^2 a'' \right) \frac{(\Delta t_n)^2}{2}
\]

\[
+ \left( a b' + \frac{1}{2} b^2 b' \right) \left( \Delta t_n \Delta \tilde{W}_{t_n} - \Delta \tilde{Z}_{t_n} \right) + b b' \frac{(\Delta \tilde{W}_{t_n})^2}{2} - \Delta t_n + b a' \Delta \tilde{Z}_{t_n}
\]

\[
+ b \left( (b')^2 + b b'' \right) \frac{\Delta \tilde{W}_{t_n}}{2} \left( \frac{(\Delta \tilde{W}_{t_n})^2}{3} - \Delta t_n \right).
\]

(9.18)
where, as before, $a$, $b$ and the derivatives of these functions are evaluated at $\bar{X}_{t_n}$.

Multi-dimensional versions of the Milstein approximation and order 1.5 strong Itô-Taylor scheme can be found in [KP92] as well as even higher order strong approximations based on Itô-Taylor expansions.

### 9.3 Explicit strong approximation

A disadvantage of the strong Itô-Taylor expansions discussed so far is that they at each time step require the evaluation of derivatives of various orders of the drift and diffusion coefficients. We shall now consider strong approximations, which, just like Runge-Kutta methods for ordinary differential equations, avoid the use of derivatives. Due to the difference between ordinary and stochastic calculus, however, heuristic generalizations of Runge-Kutta methods are of little use.

As an example, the following generalization of Heun’s scheme for ordinary differential equations

$$X_{t_{n+1}} = X_{t_n} + a(t_n, X_{t_n}) \Delta t_n + b(t_n, X_{t_n}) \Delta W_{t_n},$$

with supporting value

$$X_{t_n} = X_{t_n} + a(t_n, X_{t_n}) \Delta t_n + b(t_n, X_{t_n}) \Delta W_{t_n},$$

converges to the solution to the Stratonovich stochastic differential equation

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

and not, as desired, to the Itô version of this equation.

Nevertheless there exist modifications of strong approximations based on Itô-Taylor expansions which are derivative-free. One such is the following explicit order 1 strong scheme proposed by Platen

$$X_{t_{n+1}} = X_{t_n} + a(t_n, X_{t_n}) \Delta t_n + b(t_n, X_{t_n}) \Delta W_{t_n} + \frac{b(t_n, X_{t_n}) - b(t_n, \bar{X}_{t_n})}{\sqrt{\Delta t_n}} \left( \frac{\Delta \bar{W}_{t_n}}{2} - \Delta t_n \right),$$

with supporting value

$$\bar{X}_{t_n} = X_{t_n} + a(t_n, X_{t_n}) \Delta t_n + b(t_n, X_{t_n}) \sqrt{\Delta t_n},$$

This is a derivative-free version of the Milstein approximation. An application of the deterministic Taylor expansion easily shows that the ratio

$$\frac{b(t_n, X_{t_n}) + a(t_n, X_{t_n}) \Delta t_n + b(t_n, X_{t_n}) \sqrt{\Delta t_n} - b(t_n, \bar{X}_{t_n})}{\sqrt{\Delta t_n}}$$

is a forward difference approximation of $b \frac{\partial b}{\partial x}$ at $(t_n, X_{t_n})$. The price that must be paid for removing the derivatives is that $b$ must be evaluated one extra time for each time step.

Strong explicit schemes of orders 1.5 and 2 are found in [KP92], where it is also shown (Theorem 11.5.1) that replacing the derivatives of the Itô coefficient functions in the strong Itô-Taylor scheme of order $\gamma$ with finite difference approximations converging strongly with order $2\gamma - \phi(\alpha)$, where

$$\phi(\alpha) = \begin{cases} 
2l(\alpha) - 2, & \text{for } l(\alpha) = n(\alpha), \\
l(\alpha) + n(\alpha) - 1, & \text{for } l(\alpha) \neq n(\alpha),
\end{cases}$$

84
then the corresponding derivative-free scheme converges strongly with order $\gamma$. Hence one can construct derivative-free strong schemes of arbitrary order. Higher order derivative-free schemes, however, require that the coefficients are evaluated at a number of extra times for each time step and the number of extra evaluations needed increases fast with increasing order of convergence.

### 9.4 Implicit strong approximation

As we saw in Section 7.7 implicit strong approximations of stochastic differential equations become stochastically numerically unstable if the expression for $X_{t_{n+1}}$ contains reciprocals of Gaussian random variables, since these do not have finite moments. This can be avoided, however, by letting the terms involving $\Delta W_{t_n}$ be evaluated at $(t_n, X_{t_n})$ instead of at $(t_{n+1}, X_{t_{n+1}})$. The main benefit of using stochastically numerically stable implicit strong approximations is that they are generally not as sensitive in the case of stiff stochastic differential equations.

The simplest implicit strong approximation is the implicit Euler approximation, given by

$$X_{t_{n+1}} = X_{t_n} + a(t_{n+1}, X_{t_{n+1}}) \Delta t_n + b(t_n, X_{t_n}) \Delta W_{t_n}. \tag{9.26}$$

As a matter of fact, we can define an entire family of implicit Euler approximations

$$X_{t_{n+1}} = X_{t_n} + \left((1 - \beta) a(t_n, X_{t_n}) + \beta a(t_{n+1}, X_{t_{n+1}})\right) \Delta t_n + b(t_n, X_{t_n}) \Delta W_{t_n}, \tag{9.27}$$

where the parameter $\beta \in [0, 1]$ characterizes the degree of implicitness. $\beta = 0$ corresponds to the explicit Euler approximation and $\beta = 1$ to the implicit Euler approximation.

There are implicit versions of the Milstein approximation, higher order Itô-Taylor schemes and derivative-free Itô-Taylor schemes as well. Generally these approximations are generated by changing the evaluation point of Itô coefficient functions with multi-indices containing only zeros from $(t_n, X_{t_n})$ to $(t_{n+1}, X_{t_{n+1}})$. As an example the implicit Milstein approximation is given by

$$X_{t_{n+1}} = X_{t_n} + a(t_{n+1}, X_{t_{n+1}}) \Delta t_n + b(t_n, X_{t_n}) \Delta W_{t_n} + \frac{1}{2} b(t_n, X_{t_n}) \frac{\partial b}{\partial x} (t_n, X_{t_n}) \left((\Delta W_{t_n})^2 - \Delta t_n\right). \tag{9.28}$$

It can be shown (cf. Section 12.6 of [KP92]) that strong implicit approximations have the same order of convergence as their corresponding explicit approximations.

### 10 Weak approximation

In this chapter we study the order of convergence of weak numerical approximations of stochastic differential equations. We restrict our attention to the case of approximations of stochastic differentials with $d = m = 1$. In Section 7.6 we found that the Euler approximation converges weakly with order 1. For weak convergence it suffices that the distribution of $X$ tends to the distribution of $X$. Hence it is not necessary to use Gaussian variables in the Euler approximation and in fact it holds that the simplified weak Euler approximation given by

$$\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + a(t_n, \bar{X}_{t_n}) \Delta t_n + b(t_n, \bar{X}_{t_n}) \Delta \bar{W}_{t_n}, \tag{10.1}$$

where the $\Delta \bar{W}_{t_n}$ are $\mathcal{F}_{t_n}$-measurable random variables satisfying the moment constraints

$$|E[\Delta \bar{W}_{t_n}]| + |E[(\Delta \bar{W}_{t_n})^2] - \Delta t_n| + |E[(\Delta \bar{W}_{t_n})^3]| \leq K (\Delta t_n)^2, \tag{10.2}$$

85
still converges weakly with order 1. A very simple example of a random variable satisfying (10.2) is the two-point distributed random variable with

\[ P(\Delta W_{t_n} = \pm \sqrt{\Delta t_n}) = \frac{1}{2}. \] (10.3)

### 10.1 Higher-order weak Itô-Taylor approximation

Similarly to the case of strong approximations, higher-order weak approximations can be derived from the Itô-Taylor expansion for some appropriate choices of hierarchical sets. We have the following result, the proof of which is omitted.

**Theorem 10.1 (14.5.1 in [KP92])** Let \( X \) be the solution to an autonomous stochastic differential equation where the drift and diffusion coefficients are Lipschitz continuous and, for some positive integer \( \beta \), belong to the class \( C^{2(\beta+1)}_{\mathcal{P}} \) defined in Section 7.2. Furthermore suppose that for all \( g \in C^{2(\beta+1)}_{\mathcal{P}} \), there exists a constant \( C \) independent of \( \delta \) such that the Itô-Taylor expansion \( \bar{X} \) with hierarchical set \( \Gamma_\beta \) satisfies

\[ |E[g(X_T)] - E[g(\bar{X}_T)]| \leq C\delta^\beta. \] (10.5)

The hierarchical sets corresponding to the two lowest order weak Itô-Taylor schemes are

\[ \Gamma_1 = \{ \alpha \in \mathcal{M} : l(\alpha) \leq 1 \} = \{v,(0),(1)\}, \] (10.6)

\[ \Gamma_2 = \{ \alpha \in \mathcal{M} : l(\alpha) \leq 2 \} = \{v,(0),(1),(0,0),(1,0),(1,1),(1,1)\}. \] (10.7)

These hierarchical sets agree with the Euler approximation and the order 2 weak Itô-Taylor scheme, the latter of which can be written

\[
\bar{X}_{t_{n+1}} = \bar{X}_{t_n} + a\Delta t_n + b\Delta W_{t_n} + \left(aa' + \frac{1}{2}b^2a''\right)\left(\frac{\Delta t_n}{2}\right)^2 + \left(ab' + \frac{1}{2}b^2b''\right)(\Delta t_n\Delta W_{t_n} - \Delta Z_{t_n}) + bn\Delta Z_{t_n} + b'n(\Delta W_{t_n})^2 - \Delta t_n,
\] (10.8)

in the autonomous case. Here all functions are evaluated at \( \bar{X}_{t_n} \) and we used \( \Delta Z_{t_n} \), as before, to denote the multiple Itô integral \( I_{(1,0)} \). As we shall see below it suffices to use an approximation \( \Delta \tilde{Z}_{t_n} \) of \( \Delta Z_{t_n} \) with an incorrect value of the first moments as long as the error is of order 3 or higher. One simple choice is to use

\[ \Delta \tilde{Z}_{t_n} = \frac{1}{2}\Delta t_n\Delta W_{t_n}, \] (10.9)

which reproduces the expectation of \( \Delta Z_{t_n} \) but, by (9.15), results in the error

\[ \left|E\left[\left(\Delta \tilde{Z}_{t_n}\right)^2 - \left(\Delta Z_{t_n}\right)^2\right]\right| = \left|\frac{(\Delta t_n)^3}{4} - \frac{(\Delta t_n)^3}{3}\right| = \frac{(\Delta t_n)^3}{12}, \] (10.10)
in the variance. It is easy to see that the error in all higher moments will be of even higher order so the error arising by using $\Delta Z_{t_n}$ instead of $\Delta Z_{t_n}$ is of sufficiently high order. With this choice of $\Delta Z_{t_n}$, the order 2 weak Itô-Taylor scheme reduces to

$$X_{t_{n+1}} = X_{t_n} + a\Delta t_n + b\Delta W_{t_n} + \left(aa' + \frac{1}{2}b^2a''\right)\left(\frac{\Delta t_n}{2}\right) + \frac{1}{2}\left(ab' + a'b + \frac{1}{2}b^2b''\right)\Delta t_n \Delta W_{t_n}$$

$$+ bb'\left(\Delta W_{t_n}\right)^2 - \frac{\Delta t_n}{2},$$

(10.11)

where, as before, all functions are evaluated at $X_{t_n}$.

As mentioned above modifications of weak Itô-Taylor schemes converge weakly with the same order as the original scheme as long as the moments of the simulations of multiple Itô integrals are correct up to some specified order. Under regularity assumptions that are similar to those in Theorem 10.1, Theorem 14.5.2 in [KP92] shows that if, for some positive integer $K > 0$ and $r \in \{1, 2,..., 2\beta + 1\}$, it holds that

$$E \left(\left(X_{t_{n+1}} - X_{t_n}\right)^h - \left(\sum_{\alpha \in \Gamma_\beta \setminus \{v\}} f_v (X_{t_n}) I_\alpha \right)^h |A_{t_n}\right) \leq K \left(1 + \max_{0 \leq n \leq n(T)} |X_{t_n}|^{2r}\right) \delta^\beta \Delta t_n,$$

(10.12)

then the modified Itô-Taylor scheme $\bar{X}$ converges weakly to $X$ with order $\beta$.

For the order 2 weak Itô-Taylor scheme we have $2\beta + 1 = 5$, so in this case we need the first five moments of the multiple Itô integrals in the Itô-Taylor expansion

$$\sum_{\alpha \in \Gamma_2 \setminus \{v\}} f_\alpha (\bar{X}_{t_n}) I_\alpha$$

(10.13)

to differ with at least order three from the first five moments of the modified Itô-Taylor scheme $X_{t_{n+1}} - X_{t_n}$. We have already seen that setting $\Delta Z_{t_n}$ as in (10.9) produces an error of order three on the moments, so it suffices that the moments of the simulation of the Wiener increment satisfies

$$E \left[\left(\Delta \tilde{W}_{n}\right)^h + E \left[\left(\Delta \tilde{W}_{n}\right)^2 - \Delta t_n\right] + E \left[\left(\Delta \tilde{W}_{n}\right)^3\right]\right]$$

$$+ E \left[\left(\Delta \tilde{W}_{n}\right)^4 - 3(\Delta t_n)^2\right] + E \left[\left(\Delta \tilde{W}_{n}\right)^5\right]$$

$$\leq K (\Delta t_n)^3.$$ 

(10.14)

One possible choice of $\Delta \tilde{W}_{n}$ satisfying (10.14) is the three-point distributed random variable defined by

$$P \left(\Delta \tilde{W}_{n} = \pm \sqrt{3}\Delta t_n\right) = \frac{1}{6}, \quad P \left(\Delta \tilde{W}_{n} = 0\right) = \frac{2}{3},$$

(10.15)

which exactly reproduces the first five moments of the Wiener process and gives an error of at least order three for moments of orders exceeding five.

10.2 Explicit weak approximation

As in the case of strong approximation, there are Runge-Kutta like weak approximations which avoid the use of derivatives of the drift and diffusion coefficients. The simplest such approxima-
tion in the autonomous case is the explicit order 2 weak scheme due to Platen

\[ X_{t_{n+1}} = X_{t_n} + \frac{a(X) + a(X_{t_n})}{2} \Delta t_n + \frac{b(X^+) + b(X^-) + 2b(X_{t_n})}{4} \Delta \tilde{W}_{t_n} \]

\[ + \frac{b(X^+) - b(X^-)}{2\sqrt{\Delta t_n}} \left( \Delta \tilde{W}_{t_n} \right)^2 - \Delta t_n, \]

(10.16)

with supporting values

\[ X = X_{t_n} + a(X_{t_n}) \Delta t_n + b(X_{t_n}) \Delta \tilde{W}_{t_n}, \]

(10.17)

and

\[ X^{\pm} = X_{t_n} + a(X_{t_n}) \Delta t_n \pm b(X_{t_n}) \sqrt{\Delta t_n}. \]

(10.18)

To obtain second order, we must make sure that \( \Delta \tilde{W}_{t_n} \) satisfies the constraint in (10.14), so for example we can let \( \Delta \tilde{W}_{t_n} \) be given by (10.15). In this scheme we need to evaluate the drift coefficient at two points and the diffusion coefficient at three points per time step. Similar explicit weak schemes of higher order exist (see [KP92] for details) but they are generally complicated and require the drift and diffusion coefficients to be evaluated at a large number of extra points.

### 10.3 Implicit weak approximation and predictor-corrector methods

The simplest implicit weak approximation is the implicit Euler approximation

\[ \overline{X}_{t_{n+1}} = \overline{X}_{t_n} + a(t_{n+1}, \overline{X}_{t_{n+1}}) \Delta t_n + b(t_{n}, \overline{X}_{t_n}) \Delta \tilde{W}_{t_n}, \]

(10.19)

where \( \Delta \tilde{W}_{t_n} \) is chosen according to (10.3). This approximation converges weakly with order 1. As previously noted in the case of strong approximations, weak implicit approximations are particularly useful in the case of stiff stochastic differential equations. We can construct implicit versions of higher order weak Itô-Taylor schemes in much the same fashion as we did for strong Itô-Taylor schemes (see [KP92] for details).

The predictor-corrector methods are two-step versions of implicit schemes and are mainly used due to their numerical stability. In the first step of a predictor-corrector method we predict the value of \( X \) to be \( \tilde{X}_{t_{n+1}} \) by, for example, the Euler approximation. In the second step we correct this value by applying an implicit approximation made explicit through the use of the predicted value \( \tilde{X}_{t_{n+1}} \) instead of \( \overline{X}_{t_{n+1}} \) on the right hand side of the implicit scheme. In the autonomous case the simplest order 1 predictor-corrector method has the corrector

\[ \overline{X}_{t_{n+1}} = \overline{X}_{t_n} + a(X_{t_n}) + a(\tilde{X}_{t_{n+1}}) \Delta t_n + b(X_{t_n}) \Delta \tilde{W}_{t_n}, \]

(10.20)

and predictor

\[ \tilde{X}_{t_{n+1}} = \overline{X}_{t_n} + a(X_{t_n}) \Delta t_n + b(X_{t_n}) \Delta \tilde{W}_{t_n}, \]

(10.21)

where \( \Delta \tilde{W}_{t_n} \) is chosen according to (10.3).

We also give an example of an order 2 predictor-corrector method for the autonomous case. Here the corrector is

\[ \overline{X}_{t_{n+1}} = \overline{X}_{t_n} + a(X_{t_n}) + a(\tilde{X}_{t_{n+1}}) \Delta t_n + \Delta \Psi_{t_n}, \]

(10.22)
with
\[
\Delta \Psi_{tn} = b \Delta \tilde{W}_{tn} + bb' \left( \frac{(\Delta \tilde{W}_{tn})^2}{2} - \Delta t_n \right) + \frac{1}{2} \left( ab' + \frac{1}{2} b^2 b'' \right) \Delta t_n \Delta \tilde{W}_{tn},
\] (10.23)
and the predictor is
\[
\tilde{X}_{tn+1} = X_{tn} + a \Delta t_n + \Delta \Psi_{tn} + \frac{1}{2} a'b \Delta t_n \Delta \tilde{W}_{tn} + \left( aa'+ \frac{1}{2} b^2 a'' \right) \frac{(\Delta t_n)^2}{2},
\] (10.24)
where all functions in (10.23)-(10.24) are evaluated at $X_{tn}$. Comparing this method to (10.11), we find that the predictor exactly corresponds to the order 2 weak Itô-Taylor scheme and that the corrector corresponds to the implicit order 2 weak Itô-Taylor scheme (see [KP92] for details). Note that the implicit scheme contains less terms than its explicit counterpart. The terms that are absent in the implicit scheme reappear if we Taylor expand the $a(\tilde{X}_{tn+1})$-term in (10.22).

### 10.4 Extrapolation methods

Another way of obtaining high orders of weak convergence is to use extrapolation methods. Let $X^\delta$ be a time discrete approximation with equidistant time steps of length $\delta$ and assume that the weak order of $X$ is 1. For example we can choose $X$ to be the Euler approximation or the simplified weak Euler approximation. Use $X^\delta$ to simulate the function
\[
E \left[ g \left( X_T^\delta \right) \right],
\] (10.25)
and, in addition, repeat the simulation with time steps of length $2\delta$ and use $X^{2\delta}$ to simulate the function
\[
E \left[ g \left( X_T^{2\delta} \right) \right].
\] (10.26)
Then the following combination of (10.25)-(10.26),
\[
V_{\delta,2}^\delta = 2E \left[ g \left( X_T^\delta \right) \right] - E \left[ g \left( X_T^{2\delta} \right) \right],
\] (10.27)
is an order 2 weak extrapolation. Similarly, if $X$ is chosen to be a time discrete approximation with weak order equal to 2, then the order 4 weak extrapolation is given by
\[
V_{\delta,4}^\delta = \frac{1}{24} \left( 32E \left[ g \left( X_T^\delta \right) \right] - 12E \left[ g \left( X_T^{2\delta} \right) \right] + E \left[ g \left( X_T^{4\delta} \right) \right] \right).
\] (10.28)

In general let, for a positive integer $\beta$, $X$ be a time discrete approximation with weak order of convergence equal to $\beta$. Define the numbers
\[
0 < d_1 < d_2 < ... < d_{\beta+1} < \infty,
\] (10.29)
and, for $l = 1, 2, ..., \beta + 1$, the time step lengths $\delta_l = d_l \delta$, for some choice of $\delta > 0$. Then
\[
V_{\delta,2\beta}^\delta = \sum_{l=1}^{\beta+1} a_l E \left[ g \left( X_T^{\delta_l} \right) \right],
\] (10.30)
where
\[
\sum_{l=1}^{\beta+1} a_l = 1, \quad \sum_{l=1}^{\beta+1} a_l (d_l)^7 = 0,
\] (10.31)
for each $\gamma = \beta, \ldots, 2\beta - 1$, is an order $2\beta$ weak extrapolation. By inspection it is clear that the weak extrapolations of (10.27)-(10.28) satisfy (10.29)-(10.31). Observe that sequences of step sizes other than those proposed in (10.27)-(10.28) are possible. The orders of convergence of these extrapolations follow quite easily from the error expansions discussed in the next chapter (see Theorem 15.3.4 in [KP92] for a proof).

11 Adaptive weak approximation

All numerical approximations investigated up to now have been based on time discretizations that are given in advance and chosen independently of the stochastic differential equation whose solution we want to approximate. In contrast we devote this chapter to the study of adaptive weak approximations, where adaptive means that the time discretization is no longer fixed, but may change from problem to problem or even from trajectory to trajectory. The methods of this section are adopted from [STZ01].

We distinguish between two cases of adaptive approximations, namely those with deterministic time steps and those with stochastic time steps. In the case of deterministic time steps, we choose, for a given stochastic differential equation, a suitable time discretization and use the same time discretization in the simulation of every approximating trajectory. In the case of stochastic time steps we determine a new time discretization for every trajectory, so that the choice of time discretization depends not only on the stochastic differential equation but also on the sample path of the Wiener process.

The basis for adaptive weak approximations is the expansion of the discretization error for the Euler approximation formulated by Talay and Tubaro in 1990. It states that for the Euler approximation $X_T^{(i)}$ of the stochastic differential equation

$$
\frac{dX_t^{(i)}}{dt} = a^{(i)}(t, X_t) \, dt + b^{(i,k)}(t, X_t) \, dW^{(k)}_t,
$$

with uniform time steps $\Delta t_n = T/N$, the error arising in a weak approximation can be expanded as

$$
E \left[ g(X_T) - g(X_T^{(i)}) \right] = \int_0^T \frac{1}{N} E \left[ \frac{\partial}{\partial t} \Psi(t, X_t) \right] \, dt + O \left( \frac{1}{N^2} \right),
$$

where the function $\Psi$ is given by

$$
\Psi = \frac{1}{2} a^{(i)} a^{(j)} \frac{\partial^2 u}{\partial x_i \partial x_j} + a^{(i)} b^{(i,j)} \frac{\partial^3 u}{\partial x_i \partial x_j \partial x_v} + \frac{1}{2} \frac{\partial u}{\partial t} + \frac{1}{2} b^{(i,k)} b^{(j,k)} \frac{\partial^4 u}{\partial x_i \partial x_j \partial x_v \partial x_w},
$$

with $b^{(i,j)} = \frac{1}{2} b^{(i,k)} b^{(j,k)}$ and $u(t, x) = E^{(i)}[g(X_T)]$. Note that here, and for the rest of this chapter, we use the Einstein summation convention. The proof of this error expansion extends to non-uniform time steps as well and can also be generalized to higher-order weak Itô-Taylor schemes (see Theorem 14.6.1 in [KP92]).

The idea of adaptive weak approximations is to use the expansion (11.2)-(11.3) to refine the time steps on those time intervals where

$$
\int_{t_n}^{t_{n+1}} E \left[ \frac{\partial}{\partial t} \Psi(t, X_t) \right] \, dt,
$$

90
is larger than some given tolerance. Unfortunately the function $\Psi$ depends on the unknown solution $X$ and derivatives of the unknown function $u$. As we shall see this problem can, however, be solved with the aid of the solution to a dual problem.

### 11.1 Adaptive approximation with deterministic time steps

Theorem 2.2 in [STZ01] states that under sufficient regularity assumptions on $a, b$ and $g$, the error of the weak Euler approximation can be expanded as

$$E \left[ g (X_T) - g (\overline{X}_T) \right]$$

$$= \sum_{n=0}^{N-1} \sum_{r=1}^{M} \left( a^{(i)} (t_{n+1}, \overline{X}_{t_{n+1}} (\omega_r)) - a^{(i)} (t_n, \overline{X}_{t_n} (\omega_r)) \right) \varphi_i (t_{n+1}, \omega_r) \frac{\Delta t_n}{2M}$$

$$+ \sum_{n=0}^{N-1} \sum_{r=1}^{M} \left( d^{(i,j)} (t_{n+1}, \overline{X}_{t_{n+1}} (\omega_r)) - d^{(i,j)} (t_n, \overline{X}_{t_n} (\omega_r)) \right) \varphi_{ik} (t_{n+1}, \omega_r) \frac{\Delta t_n}{2M}$$

$$+ \sum_{n=0}^{N-1} (\Delta t_n)^2 \left( \mathcal{O} (\Delta t_n) + \sum_{k=n}^{N-1} \mathcal{O} \left( (\Delta t_k)^2 \right) \right) + \int_0^T (I_M + II_M) dt,$$

(11.5)

where $I_M$ and $II_M$ represent the statistical error of the first and second terms in (11.5), respectively. The functions $\varphi$ and $\varphi'$ appearing in (11.5) are certain dual functions defined as follows. For all $t_n, n \in \{0, ..., N - 1\}$, and $x \in \mathbb{R}^d$, we introduce

$$c^{(i)} (t_n, x) = x^{(i)} + a^{(i)} (t_n, x) \Delta t_n + \sum_{k=1}^{m} b^{(i,k)} (t_n, x) \Delta W_{t_n}^{(k)},$$

(11.6)

and note that $c^{(i)} (t_n, \overline{X}_{t_n}^{(i)}) = \overline{X}_{t_{n+1}}^{(i)}$. The spatial derivatives of $c$ quantify the sensitivity of $\overline{X}_{t_{n+1}}$ with respect to perturbations in $\overline{X}_{t_n}$ and can, hence, be used to measure the propagation of discretization error. The dual functions $\varphi_i, i \in \{1, ..., d\}$, are recursively defined by means of the backwards problem

$$\varphi_i (T, \omega_r) = \frac{\partial}{\partial x_i} g (\overline{X}_T (\omega_r)),$$

(11.7)

$$\varphi_i (t_n, \omega_r) = \frac{\partial c^{(v)}}{\partial x_i} (t_n, \overline{X}_{t_n} (\omega_r)) \varphi_v (t_{n+1}, \omega_r),$$

(11.8)

Similarly, the dual functions $\varphi_{ik}, i, k \in \{1, ..., d\}$, are recursively defined as

$$\varphi_{ik} (T, \omega_r) = \frac{\partial^2}{\partial x_i x_j} g (\overline{X}_T (\omega_r)),$$

(11.9)

$$\varphi_{ik} (t_n, \omega_r) = \frac{\partial c^{(v)}}{\partial x_i} (t_n, \overline{X}_{t_n} (\omega_r)) \frac{\partial c^{(w)}}{\partial x_j} (t_n, \overline{X}_{t_n} (\omega_r)) \varphi_{iw} (t_{n+1}, \omega_r)$$

$$+ \frac{\partial^2 c^{(v)}}{\partial x_i x_j} (t_n, \overline{X}_{t_n} (\omega_r)) \varphi_v (t_{n+1}, \omega_r),$$

(11.10)

The variances of the statistical errors $I_M$ and $II_M$ are of order $\mathcal{O} \left( (\Delta t_n)^2 \right) / M$, so it is clear that the first two terms in (11.5) are solely responsible for the weak order of convergence of the Euler
scheme being equal to 1. Consequently these two terms can be used to construct a refinement criterion for the time discretization. Note that the discretization error during each time step is estimated by a sample average over a number of trajectories, implying that an algorithm for adaptive weak approximation based on (11.5) will have deterministic time steps.

In order to simplify the adaptive algorithm with deterministic time steps, we introduce some notation. Let
\[
\rho_n (\omega_r) = \frac{1}{2 \Delta t_n} \left( \left. a^{(i)} \right| (t_{n+1}, \mathbf{X}_{t_{n+1}} (\omega_r)) - a^{(i)} \left| (t_n, \mathbf{X}_{t_n} (\omega_r)) \right. \right) \phi_i (t_{n+1}, \omega_r) + \frac{1}{2 \Delta t_n} \left( \left. d^{(i,j)} \right| (t_{n+1}, \mathbf{X}_{t_{n+1}} (\omega_r)) - d^{(i,j)} \left| (t_n, \mathbf{X}_{t_n} (\omega_r)) \right. \right) \phi'_{ij} (t_{n+1}, \omega_r) \right) \tag{11.11}
\]
Then we can write the first two terms in (11.5) compactly as
\[
\frac{1}{M} \sum_{r=1}^{M} \sum_{n=0}^{N-1} \rho_n (\omega_r) (\Delta t_n)^2. \tag{11.12}
\]
Furthermore for \( M \) independent samples \( \{Y (\omega_r)\}_{r=0}^{M} \) of a random variable \( Y \), define the sample mean \( \bar{E} [Y; M] \) and the sample standard deviation \( \bar{S} [Y; M] \) as
\[
\bar{E} [Y; M] = \frac{1}{M} \sum_{r=1}^{M} Y (\omega_r), \tag{11.13}
\]
and
\[
\bar{S} [Y; M] = \sqrt{\bar{E} [Y^2; M] - \left( \bar{E} [Y; M] \right)^2}, \tag{11.14}
\]
respectively.

The adaptive algorithm below can be divided into two halves. In the first half the time discretization is refined as long as the time discretization error and statistical time discretization error is larger than a given tolerance and in the second half the number of trajectories is increased, with the time discretization fixed, until the statistical error is reduced below some given tolerance.

**Algorithm 11.1** Choose an initial number of trajectories \( M_{(0)} \) and a coarse time discretization \( \{\Delta t_{n,(0)}\}_{n=0}^{N_{(0)}-1} \) with \( N_{(0)} \) time steps. Split the tolerance \( TOL \) into three parts
\[
TOL = TOLS + TOLT + TOLTS, \tag{11.15}
\]
representing the statistical tolerance, the time discretization tolerance and the statistical time discretization tolerance, respectively, and set \( TOLS = \frac{2}{3} TOL, TOLT = \frac{2}{9} TOL \) and \( TOLTS = \frac{1}{9} TOL \). Set \( E_{T,(0)} = 2 TOLT, E_{T,S,(0)} = 2 TOLTS, c_0 \geq 1.65, M_{CH} = 50, N_{CH} = 3 \) and \( l = 0 \). Then

while \( E_{T,(l)} + E_{T,S,(l)} > TOLT + TOLTS \) do compute \( M_{(l)} \) new trajectories of the Euler approximation with time discretization \( \{\Delta t_{n,(l)}\}_{n=0}^{N_{(l)}-1} \) and update the approximations of the time discretization error and statistical time discretization error by
\[
E_{T,(l+1)} = \left\| \bar{E} \left[ \sum_{n=0}^{N_{(l)}-1} \rho_n (\Delta t_n)^2 ; M_{(l)} \right] \right\|, \tag{11.16}
\]

92
and

$$E_{TS,(l+1)} = c_0 \frac{\tilde{S} \left[ \sum_{n=0}^{N(l)-1} \rho_n (\Delta t_n)^2 ; M(l) \right]}{\sqrt{M(l)}},$$  \hspace{1cm} (11.17)

respectively.

if $E_{TS,(l+1)} > TOL_{TS}$ then increase the number of trajectories to $M(l)$, where

$$M(l+1) = \min \left\{ \left[ \left( \frac{c_0 \tilde{S} \left[ \sum_{n=0}^{N(l)-1} \rho_n (\Delta t_n)^2 ; M(l) \right]}{0.95TOL_{TS}} \right]^2 \right] , M_{CH} \cdot M(l) \right\},$$  \hspace{1cm} (11.18)

and let the time discretization remain fixed, that is

$$\{ \Delta t_{n,(l+1)} \}_{n=0}^{N(l+1)-1} = \{ \Delta t_{n,(l)} \}_{n=0}^{N(l)-1},$$  \hspace{1cm} (11.19)

and increment $l$ by 1.

else if $E_{TS,(l+1)} > TOL$ then determine a new time discretization $\{ \Delta t_{n,(l+1)} \}_{n=0}^{N(l+1)-1}$ by dividing the time interval $[t_n, t_{n+1}]$, $n = 0, 1, \ldots, N(l)-1$, into $l_n$ uniform subintervals, where

$$l_n = \min \left\{ \max \left\{ \left| \frac{\Delta t_{n,(l)}}{\Delta t_{n,(l+1)}} \right| , 1 \right\}, N_{CH} \right\},$$  \hspace{1cm} (11.20)

and

$$\tilde{\Delta} t_{n,(l)} = \frac{TOL}{\sqrt{\int_0^1 \sqrt{E [\rho_n ; M(l)]} \, ds}}.$$  \hspace{1cm} (11.21)

Let the number of trajectories remain fixed $M_{(l+1)} = M(l)$ and increment $l$ by 1.

end if

end while

Now set $M_{S,(0)} = M(l)$, $E_{S,(0)} = 2TOL_S$ and $p = 0$. From now on let the time discretization be fixed as $\{ \Delta t_{n,(l)} \}_{n=0}^{N(l)-1}$.

while $E_{S,(p)} > TOL_S$ do compute $M_{S,(p)}$ new trajectories of the Euler approximation with time discretization $\{ \Delta t_{n,(l)} \}_{n=0}^{N(l)-1}$ and determine the sample mean

$$E_{g,(p)} = \tilde{E} \left[ g \left( \overline{X}_T \right) ; M_{S,(p)} \right],$$  \hspace{1cm} (11.22)

of the sought function. Update the statistical error

$$E_{S,(p+1)} = c_0 \frac{\tilde{S} \left[ g \left( \overline{X}_T \right) ; M_{S,(p)} \right]}{\sqrt{M_{S,(p)}}},$$  \hspace{1cm} (11.23)

and the number of trajectories

$$M_{S,(p+1)} = \min \left\{ \left[ \left( \frac{c_0 \tilde{S} \left[ g \left( \overline{X}_T \right) ; M_{S,(p)} \right]}{0.95TOL_{S}} \right]^2 \right] , M_{CH} \cdot M_{S,(p)} \right\}.$$  \hspace{1cm} (11.24)

Increment $p$ by 1.
deterministic time steps with the exception that stochastic time steps could be written as follows, using the same notation as in the case of any given trajectory at any given time step. This makes it possible to use different discretizations
\[TOL = TOL_S + TOL_T, \quad \text{where } TOL_S = \frac{2}{3}TOL\]

11.2 Adaptive approximation with stochastic time steps

The expansion in (11.5) represents the error as a sample mean over a number of simulated trajectories and cannot be used to determine the error arising from a certain trajectory. For this, we need the error expansion derived in Theorem 3.3 of [STZ01], which states that
\[
E[g(X_T) - g(\overline{X}_T)] = E \left[ \sum_{n=0}^{N-1} \tilde{\rho}_n(t_n, \overline{X}) (\Delta t_n)^2 \right] + O \left( \sqrt{\frac{TOL}{c(TOL)} \left( \frac{C(TOL)}{c(TOL)} \right)^{\frac{k_0}{2}}} \right) \sum_{n=0}^{N-1} (\Delta t_n)^2, \quad (11.25)
\]

for some \(k_0 > 16\) and
\[
\tilde{\rho}_n = \left( \frac{\partial a(i)}{\partial t} + a(j) \frac{\partial a(i)}{\partial x_j} + d(j,v) \frac{\partial^2 a(i)}{\partial x_j \partial x_v} \right) \varphi_i(t_{n+1}) \]
\[
+ \left( \frac{\partial d(i,j)}{\partial t} + a(v) \frac{\partial d(i,j)}{\partial x_v} + d(v,w) \frac{\partial^2 d(i,j)}{\partial x_v \partial x_w} + 2d(v,j) \frac{\partial a(i)}{\partial x_v} \right) \varphi_{ij}(t_{n+1}) \]
\[
+ d(v,l) \frac{\partial d(i,j)}{\partial x_v} \varphi_{ij}(t_{n+1}), \quad (11.26)
\]

Moreover, \(\varphi\) and \(\varphi'\) are the dual functions given in (11.7)-(11.10) and \(\varphi''\) is defined as
\[
\varphi''_{ij}(T) = \frac{\partial^3 g}{\partial x_i \partial x_j \partial x_l}(\overline{X}_T),
\]
\[
\varphi''_{ij}(t_n) = \frac{\partial c(v)}{\partial x_i} (t_n, \overline{X}_T) \frac{\partial c(w)}{\partial x_j} (t_n, \overline{X}_T) \frac{\partial c(z)}{\partial x_l} (t_n, \overline{X}_T) \varphi''_{vwz}(t_{n+1}) \]
\[
+ \frac{\partial^2 c(v)}{\partial x_i \partial x_j} (t_n, \overline{X}_T) \frac{\partial c(w)}{\partial x_l} (t_n, \overline{X}_T) \varphi'_{vw}(t_{n+1}) \]
\[
+ \frac{\partial c(v)}{\partial x_i} (t_n, \overline{X}_T) \frac{\partial^2 c(w)}{\partial x_j \partial x_l} (t_n, \overline{X}_T) \varphi'_{vw}(t_{n+1}) \]
\[
+ \frac{\partial^2 c(v)}{\partial x_i \partial x_l} (t_n, \overline{X}_T) \frac{\partial c(w)}{\partial x_j} (t_n, \overline{X}_T) \varphi'_{vw}(t_{n+1}) + \frac{\partial^3 c(v)}{\partial x_i \partial x_j \partial x_l} \varphi_v(t_{n+1}), \quad (11.27)
\]

In the above equations the \(\omega_r\)'s have been left out to simplify the notation. The functions \(c\) and \(C\) are some positive functions such that \(c(TOL) \leq \tilde{\rho} \leq C(TOL)\) and \(TOL/c(TOL) \to 0\) as \(TOL \to 0\).

We can use the first term on the right hand side of expansion (11.25) to estimate the error for any given trajectory at any given time step. This makes it possible to use different discretizations for different trajectories, that is to use stochastic time steps. In short, an algorithm with stochastic time steps could be written as follows, using the same notation as in the case of deterministic time steps with the exception that \(TOL = TOL_S + TOL_T\), where \(TOL_S = \frac{2}{3}TOL\)
and $TOL_T = \frac{1}{2}TOL$. Set $\delta = TOL_T/N_{(m,0)}$, where $N_{(m,0)}$ is the number of time steps in the initial coarse time discretization.

**Algorithm 11.2** Choose an initial number of trajectories $M_{(0)}$, set $l = 0$ and $E_{s,(0)} = 2TOL_S$.

```plaintext
while $E_{s,(l)} > TOL_S$ do

   for $m = 1, 2, ..., M_{(l)}$ do generate Wiener increments $\Delta W_{n,(m,0)}^{(j)}$ for the initial coarse time discretization $\{\Delta t_{n,(m,0)}\}_{n=0}^{N_{(m,0)}-1}$. Set $p = 0$ and $E_{n,(m,p)} = 2\delta$ for all $n = 0, 1, ..., N_{(m,p)} - 1$,

   while $\max_n E_{n,(m,p)} \geq \delta$ do compute $\bar{X}_{(m,p)}$ and the weight function $\tilde{\rho}_{(m,p)}$ using the known Wiener increments $\Delta W_{n,(m,p)}^{(j)}$ and set $N_{(m,p+1)} = N_{(m,p)}$. Increment $p$ by 1.

   for $n = 0, 1, ..., N_{(m,p)} - 1$ do set $E_{n,(m,p+1)} = \left| \tilde{\rho}_{n,(m,p)} (t_n, \bar{X}) \right| (\Delta t_n)^2$.

   if $E_{n,(m,p+1)} \geq \delta$ then increase $N_{(m,p+1)}$ by 1, divide the time step $\Delta t_{n,(m,p)}$ into two uniform time steps and let both these time steps belong to the refined time discretization $\{\Delta t_{n,(m,p+1)}\}_{n=0}^{N_{(m,p+1)}-1}$. Furthermore generate increments of the Wiener process on the two smaller time steps with the aid of a so-called Brownian bridge, resulting in the increments

   \[
   \Delta W_{n,(m,p)} + \zeta \quad \text{and} \quad \Delta W_{n,(m,p)} - \zeta,
   \]

   respectively, where $\Delta W_{n,(m,p)}$ was the Wiener increment of the original time step and $\zeta \in N \left( 0, \frac{(\Delta t_n)^2}{4} \right)$. Let both these Wiener increments belong to $\Delta W_{n,(m,p+1)}$.

   else do let $\Delta t_{n,(m,p)}$ belong to the time discretization $\{\Delta t_{n,(m,p+1)}\}_{n=0}^{N_{(m,p+1)}-1}$ and let the Wiener increment over this time step be $\Delta W_{n,(m,p)}$.

   end if

end for

end while

end for

Determine the sample mean $E_{g,(l)} = \tilde{E} [g (\bar{X}_{T,(l)}); M_{(l)}]$, update the statistical error

\[
E_{S,(l+1)} = c_0 \sqrt{\tilde{S} [g (\bar{X}_{T,(l)}); M_{(l)}]}.
\]

and the number of trajectories

\[
M_{(l+1)} = \min \left\{ \left[ \left( \frac{E_{S,(l+1)}}{0.95 TOL_S} \right)^2 \right]^{\frac{1}{2}} M_{CH} ; M_{(l)} \right\}.
\]

Furthermore update $\delta = TOL_T/\tilde{E} [N_{(l)}; M_{(l)}]$ and increment $l$ by 1.
end while

Accept $E_{g,(i)}$ as an approximation of $E[g(X_T)]$ since the computational error is now bounded by TOL (at least with probability $2\Phi(c_0) - 1$, where $\Phi$ is the Gaussian distribution function).

The adaptive algorithm with deterministic time steps is simpler and not as computationally demanding as the adaptive algorithm with stochastic time steps. The algorithm with stochastic time steps will, on the other hand, produce better results for problems where the setup may change from one trajectory to another according to some given distribution.

12 Variance reduction

As we saw in Section 7.3 approximations of stochastic differential equations give rise to a statistical error $\varepsilon_{\text{stat}}$, which for an increasing number $M$ of simulated trajectories is asymptotically Gaussian with mean zero and variance

$$Var(\varepsilon_{\text{stat}}) = \frac{1}{M} Var\left(g\left(\bar{X}_T\right)\right).$$ \hspace{1cm}(12.1)

In particular we found that the length of the confidence intervals for the absolute error decreased with order $M^{-\frac{1}{2}}$. In this chapter we study two methods that can be used to reduce the statistical error.

12.1 Girsanov transformation

The first method of variance reduction considered here is based on Girsanov’s theorem, which states that if we change the drift coefficient of a stochastic differential, the distribution of the process will not change dramatically. In fact, the distribution of the new process will be absolutely continuous with respect to the distribution of the original process, and the Radon-Nikodym derivatives of the probability measures can be explicitly calculated.

**Theorem 12.1 (Girsanov’s theorem)** Let $X_t \in \mathbb{R}^d$ and $\tilde{X}_t \in \mathbb{R}^d$ be two stochastic differentials given by

$$dX_t^{(i)} = a^{(i)}(t, X_t) \, dt + \sum_{k=1}^{m} b^{(i,k)}(t, X_t) \, dW_t^{(k)},$$ \hspace{1cm}(12.2)

$$d\tilde{X}_t^{(i)} = \left(a^{(i)}(t, \tilde{X}_t) - \sum_{k=1}^{m} b^{(i,k)}(t, \tilde{X}_t) f^{(k)}(t, \omega)\right) \, dt + \sum_{k=1}^{m} b^{(i,k)}(t, X_t) \, dW_t^{(k)},$$ \hspace{1cm}(12.3)

with initial values $X_0 = \tilde{X}_0 = x$. Let the functions $a : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times m}$ and $f : [0, T] \times \Omega \to \mathbb{R}^d$ satisfy the conditions of Theorem 4.4. Put

$$\tilde{W}_t^{(k)} = W_t^{(k)} - \int_0^t f^{(k)}(s, \omega) \, ds,$$ \hspace{1cm}(12.4)

and

$$M_t = \exp\left(\sum_{k=1}^{m} \int_0^t f^{(k)}(s, \omega) \, dW_s^{(k)} - \frac{1}{2} \sum_{k=1}^{m} \int_0^t \left(f^{(k)}(s, \omega)\right)^2 \, ds\right).$$ \hspace{1cm}(12.5)
Assume that $M_t$ is a martingale with respect to $\mathcal{A}^{(m)}_t$ and $P$. Define the measure $Q$ by

$$dQ(\omega) = M_T(\omega) \, dP(\omega).$$

(12.6)

Then $Q$ is a probability measure on $\mathcal{A}_t^{(m)}$, $\tilde{W}_t$ is a Wiener process with respect to $Q$ and the process $\tilde{X}_t$ is a stochastic differential in terms of $\tilde{W}_t$ with the representation

$$d\tilde{X}_t^{(i)} = a^{(i)}(t, \tilde{X}_t) \, dt + \sum_{k=1}^m b^{(i,k)}(t, \tilde{X}_t) \, d\tilde{W}_t^{(k)}.$$  

(12.7)

Furthermore the $Q$-distribution of $\tilde{X}$ is the same as the $P$-distribution of $X$.

**Sketch of the proof.** By construction $Q$ is a measure on $\mathcal{A}_t^{(m)}$. In addition it is a probability measure since

$$Q(\Omega) = \int_\Omega dQ = \int_\Omega M_T dP = \int_\Omega M_0 dP = \int_\Omega dP = P(\Omega) = 1,$$

(12.8)

where the third equality follows by the martingale property of $M_t$.

To show that $\tilde{W}_t^{(k)}$ is a Wiener process with respect to $Q$, we can use Theorem 8.6.1 in [Øks98], which roughly states that this is true if

1. $\tilde{W}_t = (\tilde{W}_t^{(1)}, \ldots, \tilde{W}_t^{(m)})$ is a martingale with respect to $Q$.
2. $\tilde{W}_t^{(k)} \tilde{W}_t^{(l)} - \delta_{kl} t$ is a martingale with respect to $Q$ for all $k, l \in \{1, 2, \ldots, m\}$.

The verification of these two properties is left as an exercise (see the proof of Theorem 8.6.4 in [Øks98]). The representation of $\tilde{X}_t$ in (12.7) follows immediately by inserting (12.4) in (12.7) and comparing the result to (12.3). Finally the last statement is a consequence of weak uniqueness (see Section 4.3), since the drift and diffusion coefficients of (12.2) and (12.7) coincide.

The aim of weak approximation is to approximate

$$u(t, x) = E^{t,x}[g(X_T)],$$

(12.9)

for some stochastic differential $X$, some real-valued function $g$ and some choice of $(t, x)$. Suppose that the stochastic differential $\tilde{X}_t^{(i)}$ and martingale $M_t$ defined in Theorem 12.1 exist. Then since the $Q$-distribution of $\tilde{X}$ and the $P$-distribution of $X$ coincide, we obtain

$$E^{t,x}[g(X_T)] = \int_\Omega g\left(X_T^{t,x}\right) \, dP = \int_\Omega g\left(\tilde{X}_T^{t,x}\right) \, dQ = \int_\Omega g\left(\tilde{X}_T^{t,x}\right) M_T \, dP = E^{t,x}\left[g\left(\tilde{X}_T\right) M_T\right],$$

(12.10)

and we can estimate the expectation of

$$g\left(\tilde{X}_T\right) M_T,$$

(12.11)

where $\tilde{X}_t = x$, to evaluate (12.9). This result is independent of the choice of functions $f^{(j)}$ in Theorem 12.1. Hence we would like to be able to find some choice of $f^{(j)}$ that reduces the variation of $g\left(\tilde{X}_T\right) M_T$ considerably.
In Section 5.2, we saw that under sufficient smoothness conditions on the function \( g \) and the drift and diffusion coefficients of \( X \), the function \( u \) satisfies the Kolmogorov backward equation

\[
\frac{\partial u}{\partial t} + \sum_{i=1}^{d} a^{(i)} \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \sum_{k=1}^{m} b^{(i,k)} b^{(j,k)} \frac{\partial^2 u}{\partial x_i \partial x_j} = 0,
\]  

(12.12)

with terminal condition

\[
u(T, x) = g(x).
\]  

(12.13)

We now show that if we suppose that \( u > 0 \) and choose the function \( f^{(j)} \) as

\[
f^{(j)} = -\frac{1}{u} \sum_{i=1}^{d} b^{(i,j)} \frac{\partial u}{\partial x_i},
\]  

(12.14)

then

\[
g \left( X^0_T \right) M_T = u \left( T, X^0_T \right) M_T = u \left( 0, x \right),
\]  

(12.15)

so with this choice (12.11) is non-random and we have successfully reduced the variance to zero. Indeed, applying Itô’s formula to the function

\[
h \left( t, X^0_t, M_t \right) = u \left( t, X^0_t \right) M_t,
\]  

(12.16)

we obtain

\[
u \left( T, X^0_T \right) M_T = u \left( 0, X^0_0 \right) M_0 + \int_0^T d \left( u \left( t, X^0_t \right) M_t \right) = u \left( 0, x \right) + \int_0^T dh \left( t, X^0_t, M_t \right),
\]  

(12.17)

so (12.15) follows if the integral in (12.17) is zero. To simplify the notation in the calculations below we suppress the superscript 0, x and use the Einstein summation convention.

\[
\frac{dh \left( t, \tilde{X}_t, M_t \right)}{dt} = \frac{\partial h}{\partial t} dt + \frac{\partial h}{\partial x_i} d\tilde{X}_t^{(i)} + \frac{\partial h}{\partial m} dM_t + \frac{1}{2} \frac{\partial^2 h}{\partial m^2} (dM_t)^2 + \frac{1}{2} \frac{\partial^2 h}{\partial x_i \partial x_j} d\tilde{X}_t^{(i)} d\tilde{X}_t^{(j)} + \frac{\partial^2 h}{\partial x_i \partial m} d\tilde{X}_t^{(i)} dM_t.
\]  

(12.18)

Applying (12.3) and using the expression for \( dM_t \) derived in Exercise 27, the right hand side (12.18) reduces to

\[
M_t \frac{\partial u}{\partial t} dt + M_t \frac{\partial u}{\partial x_i} \left( \left( a^{(i)} - b^{(i,k)} f^{(k)} \right) dt + b^{(i,k)} dW_t^{(k)} \right) + uM_t \left( f^{(k)} dW_t^{(k)} \right)
\]

\[
+ 0 + \frac{1}{2} M_t \frac{\partial^2 u}{\partial x_i \partial x_j} b^{(i,k)} b^{(j,k)} dt + \frac{\partial u}{\partial x_i} b^{(i,k)} M_t \left( f^{(l)} dW_t^{(l)} \right)
\]

\[
= M_t \left( \frac{\partial u}{\partial t} + a^{(i)} \frac{\partial u}{\partial x_i} + \frac{1}{2} b^{(i,k)} b^{(j,k)} \frac{\partial^2 u}{\partial x_i \partial x_j} \right) dt + M_t \left( \frac{\partial u}{\partial x_i} b^{(i,k)} + u f^{(k)} \right) dW_t^{(k)}.
\]  

(12.19)

The first term on the right hand side of (12.19) is zero due to (12.12) and the second term on the right hand side of (12.19) vanishes if we choose \( f^{(k)} \) according to (12.14). Consequently the integral in (12.17) is zero and we have found a way to reduce the variance to zero. However, we have used the unknown solution to the Kolmogorov backward equation in the definition of \( f^{(k)} \) so in practice the best we can hope to achieve is to make a clever guess \( \tilde{u} \) which is close to \( u \) and use \( \tilde{u} \) in the definition of \( f^{(k)} \). If the guess is sufficiently good the variance can be reduced considerably in this case as well.

98
12.2 Variance reduced estimator

In Section 2.1 we introduced the probability density $p(s, x; t, y)$ of a Markov process $X_t$, representing the density that the process evolves from location $x$ at time $s$ to location $y$ at time $t$. For Markov processes one can prove the Chapman-Kolmogorov equation

$$p(s, x; t, y) = \int_{-\infty}^{\infty} p(s, \tau, z) p(\tau, z; t, y) \, dz,$$  
(12.20)

for all $s \leq \tau \leq t$. Stochastic differentials are Markov processes and hence the probability density of a stochastic differential will, if it exists, satisfy (12.20) as well.

Consider a time discretization

$$0 = t_0 < t_1 < t_2 < \ldots < t_N = T,$$  
(12.21)

and let $\rho(dx)$ be the probability measure of the initial value $X_0$. Then we have the representation

$$E[g(X_T)] = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \prod_{i=1}^{N} p(t_{i-1}, x_{i-1}; t_i, x_i) g(x_N) \, d\rho(x_0) \, dx_1 \ldots dx_N,$$  
(12.22)

where the second equation follows from the Chapman-Kolmogorov equation. Setting

$$d\mu(\zeta) = d\rho(x_0) \, dx_1 \ldots dx_N,$$  
(12.23)

and

$$F(\zeta) = \prod_{i=1}^{N} p(t_{i-1}, x_{i-1}; t_i, x_i) g(x_N),$$  
(12.24)

where $\zeta = (x_0, \ldots, x_N) \in (\mathbb{R}^d)^{N+1} := \Gamma$, we can write (12.22) as the finite-dimensional integral

$$E[g(X_T)] = \int_{\Gamma} F(\zeta) \, d\mu(\zeta).$$  
(12.25)

Let $\zeta$ be a random variable with density $D$ with respect to the measure $d\mu$ and suppose that $D$ is nonzero for all $\zeta = (x_0, \ldots, x_N) \in \Gamma$ such that $g(x_N) \neq 0$. Then we can use the quotient

$$\eta = \frac{F(\zeta)}{D(\zeta)},$$  
(12.26)

as a Monte-Carlo estimator for $E[g(X_T)]$. This is valid since

$$E[\eta] = E\left[ \frac{F(\zeta)}{D(\zeta)} \right] = \int_{\Gamma} \frac{F(\zeta)}{D(\zeta)} D(\zeta) \, d\mu(\zeta) = \int_{\Gamma} F(\zeta) \, d\mu(\zeta) = E[g(X_T)].$$  
(12.27)

In general the function $F$ is not explicitly known. Nevertheless, the method can still be useful if we are able to approximate the density

$$Q = \prod_{i=1}^{N} p(t_{i-1}, x_{i-1}; t_i, x_i)$$  
(12.28)
of the stochastic differential with a similar, but known density. The density

$$\tilde{Q} = \prod_{i=1}^{N} \tilde{p}(t_{i-1}, x_{i-1}; t_i; x_i),$$

(12.29)
given by the Euler approximation of the stochastic differential with time discretization (12.21) is the simplest such approximate density. In this case \(\tilde{p}\) is explicitly known and equals

$$\tilde{p}(t_{i-1}, x_{i-1}; t_i; x_i) = \frac{1}{\sqrt{\text{det}(\sigma(t_{i-1}, x_{i-1}))}(2\pi\Delta t)^{n/2}} \exp\left(-\frac{(\sigma^{-1}(t_{i-1}, x_{i-1}) A)^T A}{2\Delta t}\right),$$

(12.30)

where

$$A = x_i - x_{i-1} - a(t_{i-1}, x_{i-1})\Delta t_n,$$

(12.31)

and the symmetric matrix

$$\sigma(t_{i-1}, x_{i-1}) = b(t_{i-1}, x_{i-1})b^T(t_{i-1}, x_{i-1}),$$

(12.32)
is assumed to be positive definite for all \((t_{i-1}, x_{i-1}) \in [0, T] \times \mathbb{R}^d\). By approximating \(Q\) with \(\tilde{Q}\), we obtain the variance reducing Euler estimator

$$\tilde{\eta}_E = \frac{\tilde{Q}(\zeta)}{D(\zeta)} g(x_N),$$

(12.33)

where \(\zeta = (x_0, ..., x_N) \in \Gamma\) is a random variable whose density with respect to \(d\mu = D\).

It remains to determine a good choice for the density \(D\). For the estimator in (12.26), we can show that the optimal choice of \(D\) is given by

$$D_{opt}(\zeta) = \frac{|F(\zeta)|}{\int_{\Gamma}|F(\zeta)| d\mu(\zeta)} = \frac{Q(\zeta) |g(x_N)|}{E[|g(\hat{X}_T)|]},$$

(12.34)

for \(\zeta = (x_0, ..., x_N) \in \Gamma\). Indeed, with this choice of \(D\) the variance of \(\eta\) is (to simplify the notation we have suppressed the variable \(\zeta\) in the equation below)

$$\text{Var} \left( \frac{F}{D_{opt}} \right) = E\left[ \frac{F}{D_{opt}} \right]^2 - \left( E\left[ \frac{F}{D_{opt}} \right] \right)^2 = \int_{\Gamma} \left( \frac{F}{D_{opt}} \right)^2 D_{opt} d\mu - \left( \int_{\Gamma} \frac{F}{D_{opt}} D_{opt} d\mu \right)^2$$

$$= \int_{\Gamma} \left( \int_{\Gamma} \left| \frac{F}{D_{opt}} \right| d\mu \right) d\mu - \left( \int_{\Gamma} \frac{F}{D_{opt}} d\mu \right)^2 = \left( \int_{\Gamma} |F| d\mu \right)^2 - \left( \int_{\Gamma} F d\mu \right)^2$$

(12.35)

where we used (12.34) in third equality. For any other choice of the density \(D\), Jensen’s inequality (1.39) with \(g(x) = x^2\) yields

$$\text{Var} \left( \frac{F}{D} \right) = E\left[ \frac{F}{D} \right]^2 - \left( E\left[ \frac{F}{D} \right] \right)^2 \geq \left( E\left[ \frac{F}{D} \right] \right)^2 - \left( E\left[ \frac{F}{D} \right] \right)^2$$

$$= \left( \int_{\Gamma} |F| d\mu \right)^2 - \left( \int_{\Gamma} F d\mu \right)^2,$$

(12.36)

so the minimal possible variance is obtained by choosing \(D = D_{opt}\).

The method of selecting \(D\) as close to \(D_{opt}\) as possible is known as importance sampling. In many cases the factors of \(D_{opt}\) are unknown making it hard to find a suitable \(D\). Nevertheless, in Section 16.3 of [KP92] some cases for which significant reduction of the error can be obtained through importance sampling are discussed.
References


