# 1
Fundamental Concepts

1.1
Wiener Process, Adapted Processes and Quadratic Variation

Stochastic processes represent a fundamental concept used to model the development of a physical or nonphysical system in time. It has turned out that the apparatus of stochastic processes is powerful enough to be applied to many other fields, such as economy, finance, engineering, transportation, biology and medicine.

To start with, we recall that a random variable X is a mapping $X : \Omega \to \mathbb{R}$ that assigns a real value to each elementary event $\omega \in \Omega$. The concrete value $X(\omega)$ is called a realization. It is the value we observe after the experiment has been done. To create a mathematical machine we suppose that a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ is given. Ω is the set of all elementary events and $\mathfrak{F}$ is the family of events we are interested in. It contains the set of all elementary events Ω and is assumed to be closed with respect to forming the complement and countable intersections and unions of events from this collection of events. Such families of sets or events are called σ-algebras. The character σ indicates that even the union or intersection of countably many sets belongs to $\mathfrak{F}$ as well. For mathematical reasons we have to assume that 'events generated by X', i.e. sets of the type $\{\omega : X(\omega) \in I\}$, where I is an open or closed or semi-open interval, are really events; i.e. such sets are assumed also to belong to $\mathfrak{F}$. Unfortunately the collection of all intervals of the real line is not closed with respect to the operation of union. The smallest collection of subsets of the real line that is a σ-algebra and contains all intervals is called the σ-algebra of Borel sets and will be denoted by $\mathfrak{B}$. It turns out that we have not only $\{\omega : X(\omega) \in I\} \in \mathfrak{F}$ for any interval but even $\{\omega : X(\omega) \in B\} \in \mathfrak{F}$ for every Borel set B. This fact is referred to as the $\mathfrak{F}$-measurability of X.

It turns out that for any random variable X and any continuous or monotone function g the function $Y(\omega) = g(X(\omega))$ is again a random variable. This statement remains true even if we replace g by a function from a larger class of functions, called the family of all measurable functions, to which not only the continuous functions but also the pointwise limit of continuous functions belong. This class of functions is closed with respect to 'almost all' standard manipulations with
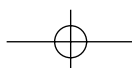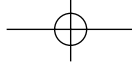
functions, such as linear combinations and products and finally forming new functions by plugging one function into another function.

The probability measure $\mathbb{P}$ is defined on $\mathfrak{F}$ and it assigns to each event $A \in \mathfrak{F}$ a number $P(A)$ called the probability of A. The mappings $A \mapsto P(A)$ satisfy the axioms of probability theory, i.e. P is a non-negative σ-additive set function on $\mathfrak{F}$ with $P(\Omega) = 1$.

We assume that the reader is familiar with probability theory at an introductory course level and in the following we use basic concepts and results without giving additional motivation or explanation.

Random variables or random vectors are useful concepts to model the random outcome of an experiment. But we have to include the additional variable 'time' when we are going to study random effects which change over time.

Definition 1.1 *By stochastic process we mean a family of random variables $(X_t)_{t \geq 0}$ which are defined on the probability space $(\Omega, \mathfrak{F}, \mathbb{P})$.*

By definition X_t is in fact a function of two variables $X_t(\omega)$. For fixed t this function of ω is a random variable. Otherwise, if we fix ω then we call the function of t defined by $t \mapsto X_t(\omega)$ a realization or a path. This means that the realization of a stochastic process is a function. Therefore stochastic processes are sometimes referred to as random functions. We call a stochastic process continuous if all realizations are continuous functions.

For the construction of a stochastic process, that is, of a suitable probability space, one needs the so-called finite dimensional distributions which are the distributions of random vectors $(X_{t_1}, \ldots, X_{t_n})$, where $t_1 < t_2 < \cdots < t_n$ is any fixed selection. For details of the construction we refer to Øksendal [175].

A fundamental idea of modeling experiments with several random outcomes in both probability theory and mathematical statistics is to start with independent random variables and to create a model by choosing suitable functions of these independent random variables. This fact explains why, in the area of stochastic processes, the particular processes with independent increments play an exceptional role. This, in combination with the fundamental meaning of the normal distribution in probability theory, makes clear the importance of the so-called Wiener process, which will now be defined.

Definition 1.2 *A stochastic process $(W_t)_{t \geq 0}$ is called a standard Wiener process or (briefly) Wiener process if:*

1) *$W_0 = 0$,*
2) *$(W_t)_{t \geq 0}$ has independent increments, i.e. $W_{t_n} - W_{t_{n-1}}, \ldots, W_{t_2} - W_{t_1}, W_{t_1}$ are independent for $t_1 < t_2 < \cdots < t_n$,*
3) *For all $0 \leq s < t$, $W_t - W_s$ has a normal distribution with expectation $\mathbb{E}(W_t - W_s) = 0$ and variance $\mathbb{V}(W_t - W_s) = t - s$,*
4) *All paths of $(W_t)_{t \geq 0}$ are continuous.*

The Wiener process is also called Brownian motion. This process is named after the biologist Robert Brown whose research dates back to the 1820s. The
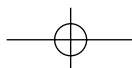
mathematical theory began with Louis Bachelier (Théorie de la Spéculation, 1900) and later by Albert Einstein (Eine neue Bestimmung der Moleküldimensionen, 1905). Norbert Wiener (1923) was the first to create a firm mathematical basis for Brownian motion.

To study properties of the paths of the Wiener process we use the quadratic variation as a measure of the smoothness of a function.

Definition 1.3 *Let $f : [0, T] \to \mathbb{R}$ be a real function and $\mathfrak{z}_n : a = t_{0,n} < t_{1,n} < \cdots < t_{n,n} = b$, a sequence of partitions with*

$$\delta(\mathfrak{z}_n) := \max_{0 \leq i \leq n-1} (t_{i+1,n} - t_{i,n}) \to 0, \quad as \quad n \to \infty.$$

If $\lim_{n\to\infty} \sum_{i=0}^{n-1} (f(t_{i+1,n}) - f(t_{i,n}))^2$ exists and is independent of the concrete sequence of partitions then this limit is called the quadratic variation of f and will be denoted by $[f]_T$.

We show that the quadratic variation of a continuously differentiable function is zero.

Lemma 1.1 *If f is differentiable in $[0, T]$ and the derivative $f'(t)$ is continuous then $[f]_T = 0$.*

Proof. Put $C = \sup_{0 \leq t \leq T} |f'(t)|$. Then $|f(t) - f(s)| \leq C|t - s|$ and

$$\sum_{i=0}^{n-1} (f(t_{i+1,n}) - f(t_{i,n}))^2 \leq C^2 \sum_{i=0}^{n-1} (t_{i+1,n} - t_{i,n})^2$$

$$\leq C^2 \delta(\mathfrak{z}_n) T \to_{n\to\infty} 0.$$

If $(X_t)_{0 \leq t \leq T}$ is a stochastic process then the quadratic variation $[X]_T$ is a random variable such that for any sequence of partitions $\mathfrak{z}_n$ with $\delta(\mathfrak{z}_n) \to 0$ it holds for $n \to \infty$

$$\sum_{i=0}^{n-1} (X_{t_{i+1,n}} - X_{t_{i,n}})^2 \to^{\mathbb{P}} [X]_T,$$

where $\to^{\mathbb{P}}$ is the symbol for stochastic convergence. Whether the quadratic variation of a stochastic process does or does not exist depends on the concrete structure of this process and has to be checked in a concrete situation and it is often more useful to deal with the convergence in mean square instead of the stochastic convergence. The relation between the two concepts provides the well known Chebyshev inequality which states that, for any random variables Z_n, Z

$$\mathbb{P}(|Z_n - Z| > \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbb{E}(Z_n - Z)^2.$$

Hence the mean square convergence $\mathbb{E}(Z_n - Z)^2 \to 0$ of Z_n to Z implies the stochastic convergence $\mathbb{P}(|Z_n - Z| > \varepsilon) \to 0$ of Z_n to Z.

Now we are going to calculate the quadratic variation of a Wiener process. To this end we need a well known fact. If V has a normal distribution with expectation μ and variance σ^2 then

$$\mathbb{E}V = \mu, \quad \mathbb{V}(V) = \mathbb{E}(V - \mu)^2 = \sigma^2$$
$$\mathbb{E}(V - \mu)^3 = 0, \quad \mathbb{E}(V - \mu)^4 = 3\sigma^4.$$

If $\mu = 0$ then

$$\mathbb{E}(V^2 - \sigma^2)^2 = \mathbb{E}(V^4 - 2\sigma^2 V^2 + \sigma^4)$$
$$= 3\sigma^4 - \sigma^4 = 2\sigma^4. \tag{1.1}$$

Theorem 1.1 *If $(W_t)_{0 \leq t \leq T}$ is a Wiener process then the quadratic variation*

$$[W]_T = T.$$

Proof. Let $\mathfrak{z}_n$ be a sequence of partitions of $[0, T]$ with $\delta(\mathfrak{z}_n) \to 0$ and put

$$Z_n = \sum_{i=0}^{n-1} (W_{t_{i+1,n}} - W_{t_{i,n}})^2.$$

From the definition of the Wiener process we get that $\mathbb{E}(W_{t_{i+1,n}} - W_{t_{i,n}})^2 = t_{i+1,n} - t_{i,n}$. As the variance of a sum of independent random variables is just the sum of the variances we get from the independent increments

$$\mathbb{E}(Z_n - t)^2 = \mathbb{E}\left(\sum_{i=0}^{n-1} (W_{t_{i+1,n}} - W_{t_{i,n}})^2 - (t_{i+1,n} - t_{i,n}) \right)^2$$

$$= \mathbb{V}(Z_n) = \sum_{i=0}^{n-1} \mathbb{V}((W_{t_{i+1,n}} - W_{t_{i,n}})^2)$$

$$= \sum_{i=0}^{n-1} \mathbb{E}((W_{t_{i+1,n}} - W_{t_{i,n}})^2 - (t_{i+1,n} - t_{i,n}))^2$$

$$= 2 \sum_{i=0}^{n-1} (t_{i+1,n} - t_{i,n})^2 \leq 2\delta(\mathfrak{z}_n)T \to 0,$$

where for the last equality we have used (1.1).

The statement $[W]_T = T$ is remarkable from different points of view. The exceptional fact is that the quadratic variation of this special stochastic process $(W_t)_{0 \leq t \leq T}$ is a degenerate random variable, it is the deterministic value T. This value is non-zero. Therefore we may conclude from Lemma 1.1 that the paths of
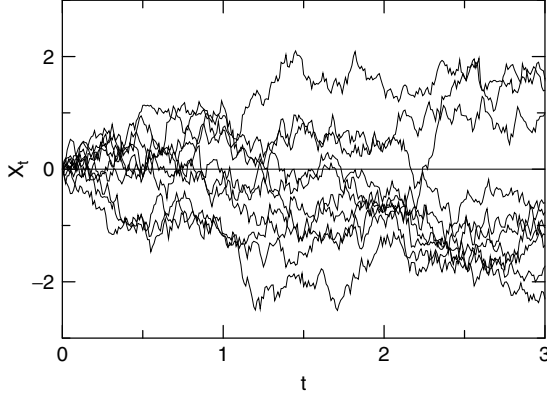
Figure 1.1 Collection of realizations X_t of the special stochastic process $(W_t)_{0 \leq t \leq T}$ named after Norbert Wiener.

a Wiener process cannot be continuously differentiable as otherwise the quadratic variation must be zero. The fact that the quadratic variation is non-zero implies that the absolute value of an increment $W_t - W_s$ cannot be proportional to $t - s$. From here we may conclude that the paths of a Wiener process are continuous but not differentiable and therefore strongly fluctuating. The illustrative picture (see Figure 1.1) of simulated realizations of a Wiener process underlines this statement.

One of the main problems in the theory of stochastic processes is to find mathematical models that describe the evolution of a system in time and can especially be used to predict, of course not without error, the values in the future with the help of information about the process collected from the past. Here and in the sequel by 'the collected information' we mean the family of all events observable up to time t. This collection of events will be denoted by $\mathfrak{F}_t$, where we suppose that $\mathfrak{F}_t$ is a σ-algebra. It is clear that $\mathfrak{F}_s \subseteq \mathfrak{F}_t \subseteq \mathfrak{F}$. Such families of σ-algebras are referred to as a *filtration* and will be denoted by $(\mathfrak{F}_t)_{\geq 0}$. Each stochastic process $(X_t)_{t \geq 0}$ generates a filtration by the requirement that $\mathfrak{F}_t$ is the smallest σ-algebra that contains all events $\{X_s \in I\}$ where I is any interval and $0 \leq s \leq t$. This filtration will be denoted $\sigma((X_s)_{0 \leq s \leq t})$. We call any stochastic process $(Y_t)_{t \geq 0}$ *adapted* to the filtration $(\mathfrak{F}_t)_{\geq 0}$ (short $\mathfrak{F}_t$-adapted) if all events that may be constructed by the process up to time t belong to the class of observable events, i.e. already belong to $\mathfrak{F}_t$. The formal mathematical condition is $\sigma((Y_s)_{0 \leq s \leq t}) \subseteq \mathfrak{F}_t$ for every $t \geq 0$. If for any fixed t and any random variable Z all events $\{Z \in I\}$, $I \subseteq \mathbb{R}$, belong to $\mathfrak{F}_t$ and it holds that $\mathbb{E}Z^2 < \infty$ then there are $X_{t_1,n}, \ldots, X_{t_{m_n},n}$, $t_{i,j} \leq t$ and (measurable) functions $f_n(X_{t_1,n}, \ldots, X_{t_{m_n},n})$ such that

$$\mathbb{E}(Z - f_n(X_{t_1,n}, \ldots, X_{t_{m_n},n}))^2 \to 0.$$

We omit the proof which would require additional results from measure theory. We denote by $\mathfrak{P}_t(X)$ the class of all such random variables. $\mathfrak{P}_t(X)$ may be considered as the *past of the process* $(X_t)_{t \geq 0}$.

Example 1.1 *Let $(W_t)_{t \geq 0}$ be a Wiener process and $\mathfrak{F}_t = \sigma((Y_s)_{0 \leq s \leq t})$.*
The following processes are $\mathfrak{F}_t$-adapted $X_t = W_t^2$, $X_t = W_{0.5 \cdot t}^2 + W_t^4$, $X_t = (W_t^4 / 1 + W_{0.1 \cdot t}^2)$. The process W_{t+1} is not $\mathfrak{F}_t$-adapted.

We fix the interval $[0, T]$, set $\mathfrak{F}_t = \sigma((W_s)_{0 \leq s \leq t})$ and denote by $\mathfrak{E}_t(W) \subseteq \mathfrak{P}_t(W)$ the collection of all elementary $\mathfrak{F}_t$-adapted processes, that is of all processes that may be written as

$$Y_t = \sum_{i=0}^{n-1} X_{t_i} I_{[t_i, t_{i+1})}(t), \quad X_{t_i} \in \mathfrak{P}_{t_i}(W), \tag{1.2}$$

where $0 = t_0 < t_1 < \cdots < t_n$ and

$$I_{[a,b)}(t) = \begin{cases} 1 & \text{if} \quad a \leq t < b, \\ 0 & \text{if} \quad \text{else.} \end{cases}$$

The $\mathfrak{F}_t$-adeptness of the process Y_t follows from the fact that exclusively random variables X_{t_i} with $t_i \leq t$ appear in the sum. The process Y_t is piecewise constant, it has the value X_{t_i} in $[t_i, t_{i+1})$ and jumps at t_i with a height

$$\Delta Y_{t_i} = X_{t_i} - X_{t_{i-1}}.$$

1.2
The Space of Square Integrable Random Variables

By $\mathcal{H}_2$ we denote the space of all random variables X with $\mathbb{E}X^2 < \infty$. Here and in the sequel we identify random variables X and Y that take on different values only with probability zero, i.e. $\mathbb{P}(X \neq Y) = 0$. Set

$$\langle X, Y \rangle := \mathbb{E}(XY).$$

It is not hard to see that $\langle X, Y \rangle$ satisfies all conditions that are imposed on a scalar product, i.e. $\langle X, Y \rangle$ is symmetric in X and Y, it is linear in both X and Y, and it holds that

$$\langle X, X \rangle \geq 0,$$

where the equality is satisfied if and only if $X = 0$.

The norm of a random variable X is given by

$$\|X\| = \sqrt{\mathbb{E}X^2},$$

and the distance of X and Y is the norm of $X - Y$. Recall that a sequence of random variables X_n is said to be convergent in mean square to X if $\mathbb{E}(X_n - X)^2 = 0$. Hence this type of convergence is nothing other than the norm convergence $\lim_{n \to \infty} \|X_n - X\| = 0$. A sequence of random variables $\{X_n\}$ is said to be a Cauchy sequence if

$$\lim_{n,m \to \infty} \|X_n - X_m\| = 0.$$

For a proof of the following theorem we refer to Øksendal [175].

Theorem 1.2 *To each Cauchy sequence $X_n \in \mathcal{H}_2$ there is some $X \in \mathcal{Z}_2$ with*

$$\lim_{n \to \infty} \|X_n - X\| = 0,$$

i.e. the space is complete.

It is clear that $\mathcal{H}_2$ is a linear space. As we have already equipped $\mathcal{H}_2$ with a scalar product we get, together with the completeness, that $\mathcal{H}_2$ is a Hilbert space. This fact allows us to apply methods from the Hilbert space theory to problems of probability theory.

A subset $\mathcal{T} \subseteq \mathcal{H}_2$ is called closed, if every limit X of a sequence $X_n \in \mathcal{T}$ belongs to $\mathcal{T}$ again. If $\mathcal{L} \subseteq \mathcal{H}_2$ is a closed linear subspace of $\mathcal{H}_2$ then there is some element in $\mathcal{L}$ that best approximates X.

Theorem 1.3 *If $\mathcal{L} \subseteq \mathcal{H}_2$ is a closed linear subspace of $\mathcal{H}_2$, then to each $X \in \mathcal{H}_2$ there is a random variable in $\mathcal{L}$, denoted by $\Pi_{\mathcal{L}} X \in \mathcal{L}$ and called the projection of X on $\mathcal{L}$, such that*

$$\inf_{Y \in \mathcal{L}} \|X - Y\| = \|X - \Pi_{\mathcal{L}} X\|.$$

Proof. Let $Y_n \in \mathcal{L}$ be a minimum sequence, i.e.

$$\lim_{n \to \infty} \|X - Y_n\| = \inf_{Y \in \mathcal{L}} \|X - Y\|.$$

Then Y_{m_n} is a minimum sequence again. Because

$$\left\| X - \frac{1}{2}(Y_n + Y_{m_n}) \right\| \leq \frac{1}{2} \|X - Y_n\| + \frac{1}{2} \left\| X - Y_{m_n} \right\|$$

$\frac{1}{2}(Y_n + Y_{m_n})$ is also a minimum sequence. Then

$$\lim_{n \to \infty} \left[\frac{1}{2} \|X - Y_n\|^2 + \frac{1}{2} \left\| X - Y_{m_n} \right\|^2 - \left\| X - \frac{1}{2}(Y_n + Y_{m_n}) \right\|^2 \right] = 0.$$

For any random variables U, V it holds that

$$\frac{1}{2} \|U\|^2 + \frac{1}{2} \|V\|^2 - \left\| \frac{1}{2}(U + V) \right\|^2 = \mathbb{E} \left(\frac{1}{2} U^2 + \frac{1}{2} V^2 - \left(\frac{1}{2}(U + V) \right)^2 \right)$$

$$= \frac{1}{4} \mathbb{E} (U - V)^2 = \frac{1}{4} \|U - V\|^2.$$

Putting $U = X - Y_n$, $V = X - Y_{m_n}$ we arrive at

$$\frac{1}{2} \|X - Y_n\|^2 + \frac{1}{2} \left\| X - Y_{m_n} \right\|^2 - \left\| X - \frac{1}{2}(Y_n + Y_{m_n}) \right\|^2$$

$$= \frac{1}{4} \left\| Y_n - Y_{m_n} \right\|^2 \to 0.$$

As m_n was an arbitrary sequence we see that Y_n is a Cauchy sequence and converges, by the completeness of $\mathcal{H}_2$, to some random variable $\Pi_{\mathcal{L}}X$ that belongs to $\mathcal{L}$ since $\mathcal{L}$ is closed by assumption.

Without going into detail we note that the projection $\Pi_{\mathcal{L}}X$ is uniquely determined in the sense that, for every $Z \in \mathcal{L}$ which also provides a best approximation, it holds that

$$\mathbb{P}(\Pi_{\mathcal{L}}X \neq Z) = 0. \tag{1.3}$$

The projection $\Pi_{\mathcal{L}}X$ can be also characterized with the help of conditions imposed on the error $X - \Pi_{\mathcal{L}}X$.

Corollary 1.1 *It holds that $Y = \Pi_{\mathcal{L}}X$ if and only if $Y \in \mathcal{L}$ and $Y - X \perp \mathcal{L}$, i.e.*

$$\langle Y - X, Z \rangle = 0 \quad \text{for every } Z \in \mathcal{L}. \tag{1.4}$$

Proof. 1. Assume $Y = \Pi_{\mathcal{L}}X$. Then $Y \in \mathcal{L}$ by the definition of the projection.

We consider

$$g(t) = \left\| (X - Y) - tZ \right\|^2 = \|X - Y\|^2 + t^2 \|Z\|^2 - 2t \langle Y - X, Z \rangle.$$

By the definition of $\Pi_{\mathcal{L}}X$ the function $g(t)$ attains its minimum at $t = 0$. Hence

$$g'(0) = -2 \langle Y - X, Z \rangle = 0$$

which implies $\langle Y - X, Z \rangle = 0$.

2. If $Y \in \mathcal{L}$ satisfies (1.4) then for every $U \in \mathcal{L}$

$$\|X - U\|^2 = \|X - Y\|^2 + 2 \langle X - Y, Y - U \rangle + \|Y - U\|^2.$$

As $Z = Y - U \in \mathcal{L}$ we see that the middle term vanishes. Hence the right-hand term is minimal if and only if $U = Y$.

The simplest prediction of a random variable X is a constant value. Which value a is the best one? It is easy to see that the function

$$\varphi(a) = \mathbb{E}(X - a)^2$$

attains the minimum at $a_0 = \mathbb{E}X$. Consequently, if $\mathcal{L}$ consists of constant random variables only, then $\Pi_{\mathcal{L}}X = \mathbb{E}X$. This is the reason why, for any closed linear subspace, we call the projection $\Pi_{\mathcal{L}}X$ the *conditional expectation* given $\mathcal{L}$. In this case we tacitly assume that all constant random variables are contained in $\mathcal{L}$. As $\mathcal{L}$ is a linear space this is equivalent to the fact that $Z_0 \equiv 1 \in \mathcal{L}$. If this condition is satisfied then we write

$$\mathbb{E}(X|\mathcal{L}) := \Pi_{\mathcal{L}}X.$$

Choosing $Z = 1$ in (1.4) we get the following.

Conclusion 1.1 *(Iterated expectation) It holds that*

$$\mathbb{E}(\mathbb{E}(X|\mathcal{L})) = \mathbb{E}X. \tag{1.5}$$

The relation (1.4) provides the orthogonal decomposition

$$X = \Pi_{\mathcal{L}}X + (X - \Pi_{\mathcal{L}}X). \tag{1.6}$$

Here $\Pi_{\mathcal{L}}X$ belongs to the subspace $\mathcal{L}$ whereas the error $X - \Pi_{\mathcal{L}}X$ is perpendicular on $\mathcal{L}$.

The Corollary 1.1 implies that the projection operator $\Pi_{\mathcal{L}}$ is linear, i.e.

$$\Pi_{\mathcal{L}}(a_1 X_1 + a_2 X_2) = a_1 \Pi_{\mathcal{L}}(X_1) + a_2 \Pi_{\mathcal{L}}(X_2).$$

The relation (1.6) implies

$$\|\Pi_{\mathcal{L}}X\| \leq \|X\|.$$

This inequality yields, in conjunction with the linearity, that $\Pi_{\mathcal{L}}X$ depends continuously on X. Indeed, $X_n \to X$ implies

$$\|\Pi_{\mathcal{L}}X_n - \Pi_{\mathcal{L}}X\| = \left\|\Pi_{\mathcal{L}}(X_n - X)\right\| \leq \|X_n - X\| \to 0. \tag{1.7}$$

Now we collect other properties of the conditional expectation that will be used in the sequel.

Lemma 1.2 *If $\mathcal{L}$ is a closed linear subspace of $\mathcal{H}_2$ that contains the constant variables and V is a random variable such that $UV \in \mathcal{L}$ for every $U \in \mathcal{L}$ then*

$$\mathbb{E}(VX|\mathcal{L}) = V\mathbb{E}(X|\mathcal{L}).$$

Proof. The assumption $VU \in \mathcal{L}$ and (1.4) imply

$$0 = \left\langle X - \mathbb{E}(X|\mathcal{L}), VU \right\rangle$$
$$= \mathbb{E}(XV - V\mathbb{E}(X|\mathcal{L}))U = \left\langle XV - V\mathbb{E}(X|\mathcal{L}), U \right\rangle.$$

The application of Corollary 1.1 completes the proof.

The multiple application of the conditional expectation corresponds to the iterated application of projections.

Lemma 1.3 *If $\mathcal{L}_i$ is a closed linear subspace of $\mathcal{H}_2$ that contains the constant variables and $\mathcal{L}_1 \subseteq \mathcal{L}_2$ then*

$$\mathbb{E}((\mathbb{E}(X|\mathcal{L}_2))|\mathcal{L}_1) = \mathbb{E}(X|\mathcal{L}_1).$$

Proof. Set $R = \mathbb{E}(X|\mathcal{L}_1)$ and $S = \mathbb{E}(X|\mathcal{L}_2)$. Then by Corollary 1.1

$$\langle X - R, U \rangle = 0 \quad \text{for every } U \in \mathcal{L}_1,$$
$$\langle X - S, U \rangle = 0 \quad \text{for every } U \in \mathcal{L}_2.$$

The assumption $\mathcal{L}_1 \subseteq \mathcal{L}_2$ gives

$$\langle S - R, U \rangle = 0 \quad \text{for every } U \in \mathcal{L}_1.$$

Corollary 1.1 completes the proof.

Next we study the relation between the independence of random variables and the conditional expectation.

Lemma 1.4 *If X is independent of every $Z \in \mathcal{L}$ then*

$$\mathbb{E}(X|\mathcal{L}) = \mathbb{E}X.$$

Proof. The required independence implies

$$\mathbb{E}(XZ) = (\mathbb{E}X)(\mathbb{E}Z)$$
$$\mathbb{E}(X - \mathbb{E}X)Z = 0.$$

The statement follows from Corollary 1.4 and the fact that the constant random variable $\mathbb{E}X$ belongs to $\mathcal{L}$.

We say that $\mathcal{L}$ is generated by the random variables $X_1, \ldots, X_n$ if $\mathcal{L}$ consists of all possible functions (not necessarily linear) $h(X_1, \ldots, X_n)$ such that $\mathbb{E}h^2(X_1, \ldots, X_n) < \infty$. Then we write

$$\mathcal{L} = \mathcal{G}(X_1, \ldots, X_n).$$

Suppose the vector $(Y, X_1, \ldots, X_n)$ has the joint density $f(y, x_1, \ldots, x_n)$. Then

$$g(x_1, \ldots, x_n) = \int f(y, x_1, \ldots, x_n) \, dy \tag{1.8}$$

is the marginal density of $(X_1, \ldots, X_n)$ and

$$f(y|x_1, \ldots, x_n) = \frac{f(y, x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)} \tag{1.9}$$

is called the conditional density of of Y given $X_1 = x_1, \ldots, X_n = x_n$.

Theorem 1.4 *Let γ be any function with $\mathbb{E}\gamma^2(Y) < \infty$ and $f(y|x_1, \ldots, x_n)$ be the conditional density of Y given $X_1 = x_1, \ldots, X_n = x_n$. Then*

$$\mathbb{E}(\gamma(Y)|\mathcal{G}(X_1, \ldots, X_n)) = \psi(X_1, \ldots, X_n),$$

where ψ is the so-called regression function that is given by

$$\psi(x_1, \ldots, x_n) = \int_{-\infty}^{+\infty} \gamma(t) f(t|x_1, \ldots, x_n) \, dt. \tag{1.10}$$

Proof. As $\mathcal{G}(X_1, \ldots, X_n)$ consists of all functions $\varphi(X_1, \ldots, X_n)$ it suffices to show that

$$\mathbb{E}(Y - \psi(X_1, \ldots, X_n))^2 \leq \mathbb{E}(Y - \varphi(X_1, \ldots, X_n))^2.$$

It holds that

$$\mathbb{E}(\gamma(Y) - \varphi(X_1, \dots, X_n))^2$$
$$= \int \cdots \int (\gamma(y) - \varphi(x_1, \dots, x_n))^2 f(y, x_1, \dots, x_n) \, dy \, dx_1 \cdots dx_n$$
$$= \int \cdots \int (\gamma(y) - \psi)^2 f(y, x_1, \dots, x_n) \, dy \, dx_1 \cdots dx_n$$
$$+ 2 \int \cdots \int (\gamma(y) - \psi)(\psi - \varphi) f(y, x_1, \dots, x_n) \, dx \, dy_1 \cdots dx_n$$
$$+ \int \cdots \int (\varphi - \psi)^2 f(y, x_1, \dots, x_n) \, dx \, dx_1 \cdots dx_n.$$

To calculate the middle term we note that $\varphi - \psi$ does not depend on y. Hence

$$\int \cdots \int (\gamma(x) - \psi)(\psi - \varphi) f(y, x_1, \dots, x_n) \, dx \, dy_1 \cdots dx_n$$
$$= \int \cdots \int \left((\psi - \varphi) \int (\gamma(y) - \psi) f(y|x_1, \dots, x_n) \, dy \right)$$
$$\times g(x_1, \dots, x_n) \, dx_1 \cdots dx_n$$
$$= 0$$

because of (1.10). Hence

$$\mathbb{E}(\gamma(X) - \varphi(X_1, \dots, X_n))^2$$
$$= \mathbb{E}(\gamma(X) - \psi(X_1, \dots, X_n))^2 + \mathbb{E}(\varphi(X_1, \dots, X_n) - \psi(X_1, \dots, X_n))^2.$$

The term on the right-hand side becomes minimal if and only if $\varphi(X_1, \dots, X_n) - \psi(X_1, \dots, X_n) = 0$ which proves the statement.

Let $(X_t)_{t \geq 0}$ be a stochastic process such that all finite dimensional distributions of $X_{t_1}, \dots, X_{t_n}$ have a density that we will denote by $f_{t_1, \dots, t_n}(x_1, \dots, x_n)$, where $t_1 < t_2 < \cdots < t_n$. By

$$f_{t_n | t_1, \dots, t_{n-1}}(x_n | x_1, \dots, x_{n-1}) = \frac{f_{t_1, \dots, t_n}(x_1, \dots, x_n)}{f_{t_1, \dots, t_{n-1}}(x_1, \dots, x_{n-1})} \tag{1.11}$$

we denote the conditional density of X_{t_n} given $X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}$. We call a stochastic process a *Markov process* if the conditional density depends only on the values of the process at the last moment of the past, i.e.

$$f_{t_n | t_1, \dots, t_{n-1}}(x_n | x_1, \dots, x_{n-1}) = f_{t_n | n-1}(x_n | x_{n-1}). \tag{1.12}$$

If $(X_t)_{t \geq 0}$ is a Markov process, then by Theorem 1.4, for every $t_1 < t_2 < \cdots < t_n = t$ and $h > 0$

$$\mathbb{E}(\gamma(X_{t+h}) | \mathcal{G}(X_{t_1}, \dots, X_{t_n})) = \mathbb{E}(\gamma(X_{t+h}) | \mathcal{G}(X_t)). \tag{1.13}$$

Conversely, if the last condition holds for every γ then

$$\int \gamma(x_n) f_{t_n|n-1}(x_n|x_{n-1})\,\mathrm{d}x_n = \int \gamma(x_n) f_{t_n|t_1,\ldots,t_{n-1}}(x_n|x_1,\ldots,x_{n-1})\,\mathrm{d}x_n, \quad (1.14)$$

As γ is arbitrary the relation (1.14) yields

$$f_{t_n|n-1}(x_n|x_{n-1}) = f_{t_n|t_1,\ldots,t_{n-1}}(x_n|x_1,\ldots,x_{n-1}).$$

Recall that $\mathfrak{P}_t(X)$ is the smallest closed subspace of $\mathcal{H}_2$ that contains all subspaces $\mathcal{G}(X_{t_1},\ldots,X_{t_m})$, where $t_1 < t_2 < \cdots < t_n$. This means that $\mathfrak{P}_t(X)$ consists of all random variables that are either functions of random variables from the past or a limit of such random variables. Hence by the continuity of the scalar product

$$\gamma(X_{t+h}) - \mathbb{E}(\gamma(X_{t+h})|\mathcal{G}(X_t)) \perp Z, \quad Z \in \mathfrak{P}_t(X),$$

if and only if

$$\gamma(X_{t+h}) - \mathbb{E}(\gamma(X_{t+h})|\mathcal{G}(X_t)) \perp Z, \quad Z \in \mathcal{G}(X_{t_1},\ldots,X_{t_n})$$

for any $t_1 < t_2 < \cdots < t_n \le t$. As $\mathbb{E}(\gamma(X_{t+h})|\mathcal{G}(X_t)) \in \mathfrak{P}_t(X)$ then from Corollary 1.1 we get the following theorem.

Theorem 1.5 *A stochastic process $(X_t)_{t \ge 0}$ is a Markov process if and only if*

$$\mathbb{E}(\gamma(X_{t+h})|\mathfrak{P}_t(X)) = \mathbb{E}(\gamma(X_{t+h})|\mathcal{G}(X_t))$$

for every function γ with $\mathbb{E}\gamma^2(X_{t+h}) < \infty$. This condition is equivalent to (1.13) for any $t_1 < t_2 < \cdots < t_n = t$.

Now we present a general construction scheme for Markov processes.

Theorem 1.6 *Let $(X_t)_{t \ge 0}$ be a stochastic process and $V(x,t,h)$ for $t,h > 0$, $x \in \mathbb{R}$ a family of random variables such that:*

1) *$V(x,t,h)$ is independent of every $Z \in \mathfrak{P}_t(X)$ for every $t,h > 0$, $x \in \mathbb{R}$*
2) *$X_{t+h} = V(X_t,t,h)$.*

Then $(X_t)_{t \ge 0}$ is a Markov process.

Proof. Assume $\mathbb{E}\gamma^2(X_{t+h}) < \infty$ and fix $t_1 < \cdots < t_n = t$. Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be the basic probability space. For fixed $t,h > 0$ the random variable $\gamma(V(x,t,h))$ is a function of x and ω, say $\Gamma(x,\omega)$. Without proof we use the fact that each such function can be approximated by linear combinations of the products of functions $v(x)V(\omega)$ in the sense that, for suitably chosen $v_{i,n}$ and $V_{i,n}$ that are independent of every $Z \in \mathfrak{P}_t(X)$

$$\mathbb{E}\left(\sum_{i=1}^{n} v_{i,n}(X_t)V_{i,n} - \gamma(V(X_t,t,h))\right)^2 \to 0.$$

In view of Theorem 1.5 and Corollary 1.1 we have to show that

$$\mathbb{E}(\gamma(X_{t+h}) - \mathbb{E}(\gamma(X_{t+h})|\mathcal{L}(X_{t_1},\ldots,X_{t_n})))Z = 0$$

for every $Z \in \mathcal{G}(X_{t_1}, \ldots, X_{t_n})$. Due to the continuity of the projection, see (1.7), it suffices to show that

$$\mathbb{E}\left(\sum_{i=1}^{n} v_{i,n}(X_t) V_{i,n} - \mathbb{E}\left(\sum_{i=1}^{n} v_{i,n}(X_t) V_{i,n} | \mathcal{G}(X_{t_1}, \ldots, X_{t_n})\right)\right) Z = 0. \qquad (1.15)$$

To this end we note that $v_{i,n}(X_t) \in \mathcal{G}(X_{t_1}, \ldots, X_{t_n})$. Hence by Lemma 1.2

$$\mathbb{E}(v_{i,n}(X_t) V_{i,n} | \mathcal{G}(X_{t_1}, \ldots, X_{t_n})) = v_{i,n}(X_t) \mathbb{E}(V_{i,n} | \mathcal{G}(X_{t_1}, \ldots, X_{t_n})).$$

Lemma 1.4 and the independence of $V_{i,n}$ of all $X_{t_1}, \ldots, X_{t_n}$ implies

$$\mathbb{E}(V_{i,n} | \mathcal{G}(X_{t_1}, \ldots, X_{t_n})) = \mathbb{E}(V_{i,n}).$$

This yields

$$\mathbb{E}(Z(\mathbb{E}(v_{i,n}(X_t) V_{i,n} | \mathcal{L}(X_{t_1}, \ldots, X_{t_n})))) = \left[\mathbb{E}(Z v_{i,n}(X_t))\right] \left[\mathbb{E}(V_{i,n})\right]. \qquad (1.16)$$

Otherwise $V_{i,n}$ is independent of $X_{t_1}, \ldots, X_{t_n}$ and therefore independent of $Z v_{i,n}(X_t)$. This yields

$$\mathbb{E}(V_{i,n} v_{i,n}(X_t) Z) = \left[\mathbb{E}(Z v_{i,n}(X_t))\right] \left[\mathbb{E}(V_{i,n})\right]. \qquad (1.17)$$

The relations (1.16) and (1.17) imply (1.15) and thus the statement.

1.3
The Ito Integral and the Ito Formula

The aim of this section is to introduce and study the concept of the Ito integral which is an integral where, instead of the classical Riemann integral, the values of the function to be integrated are not weighted according to the length of the interval from the chosen partition. Instead we weight this values by increments of a Wiener process. A first idea could be to set

$$\int_a^b X_s \, \mathrm{d} W_s := \int_a^b X_s W_s' \, \mathrm{d}s. \qquad (1.18)$$

But we know from the discussion after Theorem 1.1 that the derivative W_s' does not exist. So this fact excludes this method. Ito succeeded in constructing an integral of the above type by starting as a first step with elementary processes and in a second step by extending the integral to a larger class of processes.

Recall that by (1.2) every elementary adapted process $X \in \mathfrak{E}(W)$ can be written as

$$Y_t = \sum_{i=0}^{n-1} X_{t_i} I_{[t_i, t_{i+1})}(t), \quad X_{t_i} \in \mathfrak{P}_{t_i}(W).$$

We set

$$\int_0^T X_s \, \mathrm{d} W_s := \sum_{i=0}^{n-1} X_{t_i} (W_{t_{i+1}} - W_{t_i}).$$

A first immediate property of this integral concept is its linearity, i.e.

$$\int_0^T (c_1 X_s^{(1)} + c_2 X_s^{(2)}) \, dW_s = c_1 \int_0^T X_s^{(1)} \, dW_s + c_2 \int_0^T X_s^{(2)} \, dW_s.$$

Another property that makes Hilbert space arguments applicable is the so-called isometry property.

Theorem 1.7 *If* $X^{(1)}, X^{(2)} \in \mathfrak{E}(W)$ *then*

$$\left\langle \int_0^T X_s^{(1)} \, dW_s, \int_0^T X_s^{(2)} \, dW_s \right\rangle = \int_0^T \left\langle X_s^{(1)}, X_s^{(2)} \right\rangle \, ds. \tag{1.19}$$

Proof. A possible change to a joint refinement shows that the two elementary processes $X_s^{(1)}$ and $X_s^{(2)}$ can be represented about the same partition. Hence

$$Y_t^{(j)} = \sum_{i=0}^n X_{t_i}^{(j)} I_{[t_i, t_{i+1})}(t),$$

with some $X_{t_i}^{(j)} \in \mathfrak{P}_{t_i}(W)$. Then

$$\left\langle \int_0^T X_s^{(1)} \, dW_s, \int_0^T X_s^{(2)} \, dW_s \right\rangle$$

$$= \sum_{i,j=0}^{n-1} \mathbb{E}(X_{t_i}^{(1)} X_{t_j}^{(2)} (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j})).$$

Let $i \neq j$ and for example $t_i > t_j$. The independence of the increments implies that $W_{t_{i+1}} - W_{t_i}$ and $X_{t_i}^{(1)} X_{t_j}^{(2)} (W_{t_{j+1}} - W_{t_j})$ are independent. Consequently $\mathbb{E}(W_{t_{i+1}} - W_{t_i}) = 0$ implies that the mixed terms vanish. This yields

$$\left\langle \int_0^T X_s^{(1)} \, dW_s, \int_0^T X_s^{(2)} \, dW_s \right\rangle = \sum_{i=0}^{n-1} \mathbb{E}(X_{t_i}^{(1)} X_{t_i}^{(2)} (W_{t_{i+1}} - W_{t_i})^2).$$

Because of $X_{t_i}^{(1)} X_{t_i}^{(2)} \in \mathfrak{P}_{t_i}(W)$ this random variable from the past is independent of $(W_{t_{i+1}} - W_{t_i})^2$ which implies

$$\mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2 X_{t_i}^{(1)} X_{t_i}^{(2)}] = \mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2] \mathbb{E}[X_{t_i}^{(1)} X_{t_i}^{(2)}]$$

$$= (t_{i+1} - t_i)[\mathbb{E}(X_{t_i}^{(1)} X_{t_i}^{(2)})].$$

Hence

$$\left\langle \int_0^T X_s^{(1)} \, dW_s, \int_0^T X_s^{(2)} \, dW_s \right\rangle = \sum_{i=0}^{n-1} [\mathbb{E}(X_{t_i}^{(1)} X_{t_i}^{(2)})](t_{i+1} - t_i)$$

$$= \int_0^T \left\langle X_s^{(1)}, X_s^{(2)} \right\rangle \, ds.$$

We denote by $\mathfrak{L}_2(W)$ the set of all $\mathfrak{P}_t(W)$-adapted processes X with

$$\int_0^T \mathbb{E}X_t^2 \, dt < \infty.$$

In the sequel we use the fact that every $X \in \mathfrak{L}_2(W)$ can be approximated by elementary processes $X^{(n)} \in \mathfrak{E}(W)$ in the sense that

$$\lim_{n \to \infty} \int_0^T \mathbb{E}(X_t^{(n)} - X_t)^2 \, dt = 0. \tag{1.20}$$

We refer to Øksendal [175] for a proof.
The relation (1.20) provides

$$\lim_{n,m \to \infty} \int_0^T \mathbb{E}(X_t^{(n)} - X_t^{(m)})^2 \, dt = 0,$$

which, together with the isometry property (1.19), leads to

$$\lim_{n,m \to \infty} \mathbb{E}\left(\int_0^T X_t^{(n)} \, dW_t - \int_0^T X_t^{(m)} \, dW_t\right)^2$$

$$= \lim_{n,m \to \infty} \int_0^T \mathbb{E}(X_t^{(n)} - X_t^{(m)})^2 \, dt = 0.$$

This means that the sequence of random variables $\int_0^T X_t^{(n)} \, dW_t$ is a Cauchy sequence and converges therefore to a random variable that will be denoted by

$$\int_0^T X_t \, dW_t.$$

This random variable is independent of the choice of the approximating sequence $X_t^{(n)}$ and is called the *Ito integral*. The continuity of the scalar product shows that the above isometry property is still valid for the larger class of processes $X \in \mathfrak{L}_2(W)$.

Theorem 1.8 *If $X, Y \in \mathfrak{L}_2(W)$ then*

$$\int_0^T (aX_t + bY_t) \, dW_t = a \int_0^T X_t \, dW_t + b \int_0^T Y_t \, dW_t$$

$$\left\langle \int_0^T X_t \, dW_t, \int_0^T Y_t \, dW_t \right\rangle = \int_0^T \langle X_t, Y_t \rangle \, dt.$$

Letting the upper bound in the integral be variable we may introduce the new stochastic process $\int_0^t X_s \, dW_s$ which has been constructed exclusively with the help of random variables from $\mathfrak{P}_t(W)$. Thus we see that the new process

$$Y_t = \int_0^t X_s \, dW_s \tag{1.21}$$

again belongs to $\mathfrak{L}_2(W)$. This process has an important projection property.

Theorem 1.9 *If $t_1 < t_2$ then Y_t in (1.21) satisfies*

$$\mathbb{E}(Y_{t_2}|\mathfrak{P}_{t_1}(W)) = Y_{t_1} \tag{1.22}$$

$$\mathbb{E}Y_{t_1} = \mathbb{E}Y_{t_2} = 0. \tag{1.23}$$

Proof. By the linearity of the Ito integral and the continuity of the projection we have to prove the statement only for elementary processes of the type $X_t = ZI_{[a,b)}(t)$ where $Z \in \mathfrak{P}_a(W)$. Then

$$Y_t = \int_0^t X_s \, \mathrm{d}W_s = Z(W_{b \wedge t} - W_a),$$

where $b \wedge t = \min(b, t)$. This shows that Y_t does not depend on t for $t < a$ and $t > b$. Hence we have only to consider the case $a \le t_1 < t_2 \le b$. Then $Y_{t_2} - Y_{t_1} = Z(W_{t_2} - W_{t_1})$ and

$$\mathbb{E}(Y_{t_2} - Y_{t_1}|\mathfrak{P}_{t_1}(W)) = \mathbb{E}(Z(W_{t_2} - W_{t_1})|\mathfrak{P}_{t_1}(W)).$$

As $Z \in \mathfrak{P}_a(W) \subseteq \mathfrak{P}_{t_1}(W)$ we may apply Lemma 1.2 and can take Z out of the conditional expectation

$$\mathbb{E}(Z(W_{t_2} - W_{t_1})|\mathfrak{P}_{t_1}(W)) = Z\mathbb{E}((W_{t_2} - W_{t_1})|\mathfrak{P}_{t_1}(W)).$$

The independence of $W_{t_2} - W_{t_1}$ and the random variables from $\mathfrak{P}_{t_1}(W)$ together with Lemma 1.4 yield

$$\mathbb{E}((W_{t_2} - W_{t_1})|\mathfrak{P}_{t_1}(W)) = 0$$

and therefore

$$\mathbb{E}(Y_{t_2} - Y_{t_1}|\mathfrak{P}_{t_1}(W)) = 0.$$

Because of $Y_{t_1} \in \mathfrak{P}_{t_1}(W)$ we obtain $\mathbb{E}(Y_{t_2}|\mathfrak{V}_{t_1}(W)) = Y_{t_1}$ which is the first statement. The relation (1.5) implies $\mathbb{E}Y_{t_2} = \mathbb{E}Y_{t_1}$ for every $0 \le t_1 \le t_2$. As $Y_0 = 0$ we get (1.23). $\square$

Stochastic processes that satisfy (1.22) are called martingales in probability theory.

Now we introduce a class of processes that turns out to be useful in order to model the evolution of a time-dependent phenomenon. A stochastic process X is called an *Ito process*, if

$$X_t = X_0 + \int_0^t A_s \, \mathrm{d}s + \int_0^t B_s \, \mathrm{d}W_s, \tag{1.24}$$

where $A, B \in \mathfrak{L}_2(W)$. It is not hard to show that the quadratic variation of $\int_0^t A_s \, \mathrm{d}s$ is zero, so that $\int_0^t A_s \, \mathrm{d}s$ is a smooth part of X_t that plays the role of a drift. The second component $\int_0^t B_s \, \mathrm{d}W_s$ is irregular as the quadratic variation is

$$[X]_t = \int_0^t B_s^2 \, \mathrm{d}s \tag{1.25}$$

which can be easily shown and does not vanish. We also write

$$\mathrm{d}X_s = A_s \, \mathrm{d}s + B_s \, \mathrm{d}W_s. \tag{1.26}$$

instead of (1.24). Ito processes admit the following interpretation. For fixed $h > 0$ the increment $X_{t+h} - X_t$ is approximately given by

$$X_{t+h} - X_t \approx A_t h + B_t(W_{t+h} - W_t). \tag{1.27}$$

The first term $A_t h$ is a drift with a slope which is governed by values from the past. The factors in the product $B_t(W_{t+h} - W_t)$ are independent where $W_{t+h} - W_t$ is normally distributed with expectation zero and variance h. If the values in the past are fixed then $B_t(W_{t+h} - W_t)$ has the variance $B_t^2 h$. This mean that $B_s \, \mathrm{d}W_s$ is a diffusion term.

Diffusion processes are special Ito processes. They are characterized by the fact that the drift coefficient A_t as well as the diffusion coefficient B_t only depend on the last state of the process. This means that

$$A_t = a(t, X_t), \quad \text{and} \quad B_t = b(t, X_t),$$

with some $a(t, x)$ and $b(t, x)$. Hence

$$X_t = X_0 + \int_0^t a(s, X_s) \, \mathrm{d}s + \int_0^t b(s, X_s) \, \mathrm{d}W_s. \tag{1.28}$$

This is an integral equation for X_t, which can formally be written as a differential equation, often used as a basic equation of motion in physics and named after Langevin

$$\dot{X}_t = a(t, X_t) + b(t, X_t) \, \dot{W}_t. \tag{1.29}$$

The problem is that $\dot{W}_t \equiv \mathrm{d}W_t / \mathrm{d}t$ does not exist as we have already pointed out by showing that the paths of W_t are not differentiable.

The representation (1.28) raises the question of for which a, b the integral equation has a solution and under which conditions this solution is unique. In the sense of an initial value problem the value X_0 has to be fixed. Necessary and sufficient conditions that guarantee the existence and uniqueness of a solution of this initial value problem can be found in many books, e.g. [30, 57, 91, 104, 175].

Often the starting point X_0 is a deterministic value, say x_0. To indicate the dependence on x_0 we denote the corresponding process by X_{t, x_0}. Hence

$$X_{t, x_0} = x_0 + \int_0^t a(s, X_{s, x}) \, \mathrm{d}s + \int_0^t b(s, X_{s, x}) \, \mathrm{d}W_s, \tag{1.30}$$

and

$$X_{t+h, x} - X_{t, x} = \int_t^{t+h} a(s, X_{s, x}) \, \mathrm{d}s + \int_t^{t+h} b(s, X_{s, x}) \, \mathrm{d}W_s.$$

For every fixed x the random variable

$$V(x, t, h) = x + \int_t^{t+h} a(s, X_{s,x}) \, ds + \int_t^{t+h} b(s, X_{s,x}) \, dW_s$$

is independent of of the random variables from $\mathfrak{P}_t(W)$. If (1.30) has a unique solution then

$$X_{t+h,x_0} = V(X_{t,x_0}, t, h).$$

From Theorem 1.6 we get the Markov property.

Theorem 1.10 *If the equation*

$$X_{t,x} = x + \int_s^t a(\tau, X_{\tau,x}) \, d\tau + \int_s^t b(\tau, X_{\tau,x}) \, dW_\tau$$

has a unique solution for every x and s then the process starting at x_0 being defined as the solution of

$$X_{t,x_0} = x_0 + \int_0^t a(s, X_{s,x_0}) \, ds + \int_0^t b(s, X_{s,x_0}) \, dW_s$$

is a Markov process. It is called homogeneous, if a and b are independent of s, hence

$$X_{t,x_0} = x_0 + \int_0^t a(X_{s,x_0}) \, ds + \int_0^t b(X_{s,x_0}) \, dW_s.$$

The class of Ito processes is closed with respect to the application of smooth functions, i.e. $u(t, X_t)$ is again a Ito process whose drift and diffusion coefficient can be given explicitly.

Theorem 1.11 *(Ito formula) Suppose $A, B \in \mathcal{L}_2(W)$ and assume*

$$dX_s = A_s \, ds + B_s \, dW_s.$$

If $u : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable then

$$du(t, X_t) = \frac{\partial u}{\partial t}(t, X_t) \, dt + \frac{\partial u}{\partial x}(t, X_t) \, dX_t + \frac{1}{2} \frac{\partial^2 u}{\partial x^2}(t, X_t) \cdot (dX_t)^2, \tag{1.31}$$

where $(dX_t)^2 = dX_t \cdot dX_t$ is to be calculated according to the following rules

$$dt \cdot dt = dt \cdot dW_t = dW_t \cdot dt = 0, \tag{1.32}$$

$$dW_t \cdot dW_t = dt. \tag{1.33}$$

Proof. We give only a sketch of the proof. Further details can be found in Øksendal [175] or many other textbooks on stochastic differential equations such as Chorin and Held [30] and Karatzas and Shreve [91].

Suppose $\mathfrak{z}_n = \{t_{0,n}, \ldots, t_{n.n}\}$, $t_{0,n} = 0$, $t_{n.n} = t$ is a sequence of partitions of $[0, t]$ with $\delta(\mathfrak{z}_n) \to 0$. Then

$$u(t, X_t) - u(t, X_0) = \sum_{l=1}^{n} [u(t_{t_{l,n}}, X_{t_{l,n}}) - u(t_{t_{l-1,n}}, X_{t_{l-1,n}})].$$

and by the Taylor expansion

$$\begin{aligned}
u(t, X_{t_{l,n}}) - u(t, X_{t_{l-1,n}}) &= \frac{\partial u}{\partial t}(t_{l,n}, X_{t_{l,n}})(t_{l,n} - t_{l-1,n}) \\
&\quad + \frac{\partial u}{\partial x}(t_{l,n}, X_{t_{l,n}})(X_{t_{l,n}} - X_{t_{l-1,n}}) \\
&\quad + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t_{l,n}, X_{t_{l,n}})(X_{t_{l,n}} - X_{t_{l-1,n}})^2 + R_{l,n},
\end{aligned}$$

where

$$\sum_{l=1}^{n} R_{l,n} \overset{\mathbb{P}}{\longrightarrow} 0$$

can be shown. The sum of the first terms of the above decomposition can be shown to tend to

$$\int_0^t \frac{\partial u}{\partial t}(s, X_s)\, ds,$$

as $n \to \infty$. Similarly, by $dX_s = A_s\, ds + B_s\, dW_s$ the sum of the second terms tends to

$$\int_0^t \frac{\partial u}{\partial x}(s, X_s)A_s\, ds + \int_0^t \frac{\partial u}{\partial x}(s, X_s)B_s\, dW_s.$$

Using $dX_s = A_s\, ds + B_s\, dW_s$ again we see that the sum of the third terms consists of three parts. The first one is

$$\sum_{l=1}^{n} \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(t_{l,n}, X_{t_{l,n}})A_{t_{l-1,n}}^2 (t_{l,n} - t_{l-1,n})^2. \tag{1.34}$$

Assuming, for simplicity, a boundedness of $\frac{\partial^2 u(t,x)}{\partial x^2}A_t^2$, this sum does not exceed

$$c\sum_{l=1}^{n}(t_{l,n} - t_{l-1,n})^2 \le c\delta(\mathfrak{z}_n) \cdot t \to 0$$

as $\delta(\mathfrak{z}_n) = \max_{1 \le l \le n}|t_{l,n} - t_{l-1,n}| \to 0$. Hence (1.34) tends stochastically to zero. The second part is the mixed term

$$\sum_{l=1}^{n} \frac{\partial^2 u}{\partial x^2}(t_{l,n}, X_{t_{l,n}})A_{t_{l-1,n}}B_{t_{l-1,n}}(t_{l,n} - t_{l-1,n})(W_{t_{l,n}} - W_{t_{l,n}}).$$

If $\frac{\partial^2 u}{\partial x^2}(t_{l,n}, X_{t_{l,n}})A_{t_{l-1,n}}B_{t_{l-1,n}}$ is bounded then the expectation of the absolute value can be estimated by

$$c \sum_{l=1}^{n} (t_{l,n} - t_{l-1,n}) \mathbb{E} | W_{t_{l,n}} - W_{t_{l,n}} |.$$

Using the inequality $\mathbb{E}|Z| \le (\mathbb{E}Z^2)^{1/2}$ valid for any random variable Z we get the bound

$$c \sum_{l=1}^{n} (t_{l,n} - t_{l-1,n})(t_{l,n} - t_{l-1,n})^{1/2} \to 0$$

where we used $\max_{1 \le l \le n} |t_{l,n} - t_{l-1,n}| \to 0$ again. The sum over the third parts

$$\frac{1}{2} \sum_{l=1}^{n} \frac{\partial^2 u}{\partial x^2}(t_{l,n}, X_{t_{l,n}}) B_{t_{l-1,n}}^2 (W_{t_{l,n}} - W_{t_{l,n}})^2$$

does not disappear. By similar arguments that have been used while studying the quadratic variation of the Wiener process one can show that the last sum tends to

$$\frac{1}{2} \int_0^t \frac{\partial^2 u}{\partial x^2}(s, X_s) B_s^2 \, ds,$$

which completes the sketch of the proof.

We now consider special cases. Suppose X_t is a diffusion process already defined by (1.28)

$$X_t = X_0 + \int_0^t a(s, X_s) \, ds + \int_0^t b(s, X_s) \, dW_s \tag{1.35}$$

The transformation rules (1.32) and (1.33) give

$$
\begin{aligned}
u(t, X_t) = u(0, X_0) \\
+ \int_0^t \left[\frac{\partial u(s, X_s)}{\partial s} \, ds + \frac{\partial u(s, X_s)}{\partial x} a(s, X_s) + \frac{1}{2} \frac{\partial^2 u(s, X_s)}{\partial x^2} b^2(s, X_s) \right] ds \\
+ \int_0^t \frac{\partial u(s, X_s)}{\partial x} b(s, X_s) \, dW_s.
\end{aligned}
\tag{1.36}
$$

If u depends only on x then

$$
\begin{aligned}
u(X_t) = u(X_0) \\
+ \int_0^t \left[u'(X_s) a(s, X_s) + \frac{1}{2} u''(X_s) b^2(s, X_s) \right] ds \\
+ \int_0^t u'(X_s) b(s, X_s) \, dW_s.
\end{aligned}
\tag{1.37}
$$

Corollary 1.2 *If X_t is a solution of $dX_t = a(t, X_t) \, dt + b(t, X_t) \, dW_t$ then*

$$\mathbb{E}(u(X_t) - u(X_0)) = \mathbb{E} \int_0^t \left[u'(X_s) a(s, X_s) + \frac{1}{2} u''(X_s) b^2(s, X_s) \right] ds.$$

Proof. Theorem 1.9 shows that

$$\mathbb{E}\int_0^t \frac{\partial u(X_s)}{\partial x} b(s, X_s)\,\mathrm{d}W_s$$

is independent of t and is therefore zero as the expression vanishes for $t = 0$.

To conclude this section we note that the diffusion process X_t in (1.28) reduces to the Wiener process in the special case $a = 0, b = 1$. But in the general case one may replace the probability measure $\mathbb{P}$ by another distribution Q (Girsanov transformation) such that the process X_t becomes a Wiener process with respect to Q.

1.4
The Kolmogorov Differential Equation and the Fokker–Planck Equation

We consider the diffusion process defined by the stochastic differential equation

$$\mathrm{d}X_t = a(X_t)\,\mathrm{d}t + b(X_t)\,\mathrm{d}W_t. \tag{1.38}$$

We know from Theorem 1.10 that this process is a Markov process. As both a and b do not depend on t the process is homogeneous. Let $f(t, x, y)$ be the family of transition densities, i.e. $f(t, x, \cdot)$ is the conditional density of X_t given $X_0 = x$. If the process $X_{t,x}$ starts in $t = 0$ at x then $f(t, x, \cdot)$ is the probability density of $X_{t,x}$. This family of densities satisfies the Chapman–Kolmogorov equation

$$f(s + t, x, y) = \int f(s, x, z) f(t, z, y)\,\mathrm{d}z, \quad 0 \le s, t. \tag{1.39}$$

Let $\mathbb{C}_b$ be the space of all bounded and measurable functions $\mathbb{R}$ and denote by $\mathbb{C}_0^2$ the space of all twice continuously differentiable functions that vanish outside of some finite interval that may depend on the concrete function under consideration. For $u \in \mathbb{C}_b$ we set

$$
\begin{aligned}
(T_t u)(x) &= \int u(y) f(t, x, y)\,\mathrm{d}y \\
&= \mathbb{E}u(X_{t,x}).
\end{aligned}
$$

It is easy to see that $T_t u \in \mathbb{C}_b$. The Chapman–Kolmogorov equation implies the semigroup property, that is,

$$T_t T_s = T_{s+t}. \tag{1.40}$$

Putting $X_0 = x$ in Corollary 1.2 we get, for any $u \in \mathbb{C}_0^2$,

$$(T_t u)(x) = u(x) + \mathbb{E}\int_0^t \left[u'(X_{s,x}) a(X_{s,x}) + \frac{1}{2} u''(X_{s,x}) b^2(X_{s,x}) \right]\,\mathrm{d}s$$

and therefore

$$\frac{(T_h u)(x) - u(x)}{h} = \mathbb{E}\frac{1}{h}\int_0^h \left[u'(X_{s,x})a(X_{s,x}) + \frac{1}{2}u''(X_{s,x})b^2(X_{s,x}) \right] \mathrm{d}s.$$

Each diffusion process can be shown to be continuous. Hence $\lim_{s\downarrow 0} X_{s,x} = x$ and

$$\lim_{h\downarrow 0}\frac{1}{h}\int_0^h \left[u'(X_{s,x})a(X_{s,x}) + \frac{1}{2}u''(X_{s,x})b^2(X_{s,x}) \right]$$

$$= a(x)\frac{\partial u(x)}{\partial x} + \frac{1}{2}b^2(x)\frac{\partial^2 u(x)}{\partial x^2} = (Au)(x),$$

where A is the differential operator

$$A = a(x)\frac{\partial}{\partial x} + \frac{1}{2}b^2(x)\frac{\partial^2}{\partial x^2}. \tag{1.41}$$

This differential operator is the infinitesimal operator of the semigroup in the sense that

$$(Au)(x) = \lim_{h\downarrow 0}\frac{(T_h u)(x) - u(x)}{h}.$$

Let I be the identical operator. Then we obtain from the semigroup property (1.40) that

$$\lim_{h\downarrow 0}\frac{T_{t+h}u - T_t u}{h} = \lim_{h\downarrow 0} T_t\left(\frac{(T_h - I)u}{h}\right)$$

$$= T_t Au. \tag{1.42}$$

Similarly,

$$\lim_{h\downarrow 0}\frac{T_{t+h}u - T_t u}{h} = \lim_{h\downarrow 0}\left(\frac{(T_h - I)}{h}\right) T_t u$$

$$= AT_t u. \tag{1.43}$$

Thus we have obtained the following result.

Theorem 1.12 *If $X_{t,x}$ is the solution of*

$$\mathrm{d}X_{t,x} = a(X_{t,x})\,\mathrm{d}t + b(X_{t,x})\,\mathrm{d}W_t,$$

$$X_{0,x} = x$$

and $u \in \mathbb{C}_0^2$, then

$$u(t,x) = (T_t u)(x)$$

$$= \mathbb{E}u(X_{t,x}) = \int u(y)f(t,x,y)\,\mathrm{d}y$$

satisfies the Kolmogorov forward equation

$$\frac{\partial u(t, x)}{\partial t} = (T_t A u)(x)$$

$$= \int \left[a(\gamma) \frac{\partial u(\gamma)}{\partial \gamma} + \frac{1}{2} b^2(\gamma) \frac{\partial^2 u(\gamma)}{\partial \gamma^2} \right] f(t, x, \gamma) \, d\gamma \tag{1.44}$$

and the Kolmogorov backward equation

$$\frac{\partial u(t, x)}{\partial t} = A(T_t u)(x) \tag{1.45}$$

$$= a(x) \frac{\partial u(t, x)}{\partial x} + \frac{1}{2} b^2(x) \frac{\partial^2 u(t, x)}{\partial x^2}.$$

Proof. The statement (1.44) follows from (1.42). Similarly, (1.45) follows from (1.43). ∎

Now we establish differential equations for the transition densities. To this end we apply integration by parts. If $u, v \in \mathbb{C}_0^2$, and both a and b are twice continuously differentiable then

$$\int \left[a(x) \frac{du(x)}{dx} \right] \left[v(x) \right] dx = -\int \left[\frac{d(a(x)v(x))}{dx} \right] \left[u(x) \right] dx,$$

$$\int \left[b^2(x) \frac{d^2 u(x)}{dx^2} \right] \left[v(x) \right] dx = \int \left[\frac{d^2(b^2(x)v(x))}{dx^2} \right] \left[u(x) \right] dx.$$

The application to (1.44) yields

$$\frac{\partial u(t, x)}{\partial t} = \int \left[a(\gamma) \frac{\partial u(\gamma)}{\partial \gamma} + \frac{1}{2} b^2(\gamma) \frac{\partial^2 u(\gamma)}{\partial \gamma^2} \right] f(t, x, \gamma) \, d\gamma$$

$$= \int \left[-\frac{\partial (a(\gamma)f(t, x, \gamma))}{\partial \gamma} + \frac{1}{2} \frac{\partial^2(b^2(\gamma)f(t, x, \gamma))}{\partial \gamma^2} \right] u(\gamma) \, d\gamma.$$

Otherwise

$$\frac{\partial u(t, x)}{\partial t} = \frac{\partial}{\partial t} \int u(\gamma) f(t, x, \gamma) \, d\gamma$$

$$= \int u(\gamma) \frac{\partial}{\partial t} f(t, x, \gamma) \, d\gamma. \tag{1.46}$$

Hence for every $u \in \mathbb{C}_0^2$

$$\int \left[\frac{\partial}{\partial t} f(t, x, \gamma) + \frac{\partial(a(\gamma)f(t, x, \gamma))}{\partial \gamma} - \frac{1}{2} \frac{\partial^2(b^2(\gamma)f(t, x, \gamma))}{\partial \gamma^2} \right] u(\gamma) \, d\gamma = 0. \tag{1.47}$$

Let $u \in \mathbb{C}_0^2$ be any probability density with support, e.g.

$$u(t) = ct^2(1 - t)^2,$$

where c is determined by

$$\int_0^1 u(t)\, \mathrm{d}t = 1.$$

Put for every fixed z

$$u_n(t) = nu(n(t - z)). \tag{1.48}$$

For large n the sequence $u_n(t)$ is concentrated around z. When ψ is twice continuously differentiable we get

$$
\begin{aligned}
\int \psi(t) u_n(t)\, \mathrm{d}t &= \int \psi(t) nu(n(t - z))\, \mathrm{d}t \\
&= \int \psi\left(z + \frac{s}{n}\right) u(s)\, \mathrm{d}s \to \int \psi(z) u(s)\, \mathrm{d}s = \psi(z).
\end{aligned}
$$

The application of this statement to (1.47) yields the so-called *forward Fokker–Planck equation*

$$\frac{\partial}{\partial t} f(t, x, z) = -\frac{\partial(a(z) f(t, x, z))}{\partial z} + \frac{1}{2}\frac{\partial^2 (b^2(z) f(t, x, z))}{\partial z^2}. \tag{1.49}$$

Similarly, the relation (1.45) yields

$$
\begin{aligned}
\frac{\partial u(t, x)}{\partial t} &= a(y)\frac{\partial u(t, x)}{\partial x} + \frac{1}{2} b^2(x)\frac{\partial^2 u(t, x)}{\partial x^2} \\
&= a(y)\frac{\partial}{\partial x} \int u(y) f(t, x, y)\, \mathrm{d}y + \frac{1}{2} b^2(x)\frac{\partial^2}{\partial x^2} \int u(y) f(t, x, y)\, \mathrm{d}y \\
&= \int u(y)\left[a(y)\frac{\partial f(t, x, y)}{\partial x} + \frac{1}{2} b^2(x)\frac{\partial^2 f(t, x, y)}{\partial x^2}\right] \mathrm{d}y.
\end{aligned}
$$

Because of (1.46) we arrive at

$$\int u(y)\left[\frac{\partial}{\partial t} f(t, x, y) - a(y)\frac{\partial f(t, x, y)}{\partial x} - \frac{1}{2} b^2(x)\frac{\partial^2 f(t, x, y)}{\partial x^2}\right] \mathrm{d}y = 0.$$

Again by plugging in u_n from (1.48) and by letting $n \to \infty$ we obtain

$$\frac{\partial}{\partial t} f(t, x, y) = a(x)\frac{\partial f(t, x, y)}{\partial x} + \frac{1}{2} b^2(x)\frac{\partial^2 f(t, x, y)}{\partial x^2}, \tag{1.50}$$

which is called the *backward Fokker–Planck equation*.

1.5
Special Diffusion Processes

This section is aimed at presenting special examples of diffusion processes and studying the relation between them.

Example 1.2 *If $X_t = W_t$ is the Wiener process then $a = 0$ and $b = 1$ in the stochastic differential equation (1.38). Since $f(t, x, y)$ is the density of $x + W_t$ the family of transition densities is given by*

$$f(t, x, y) = \varphi_{0,t}(y - x)$$

where φ_{μ,σ^2} is the density of the normal distribution with parameters μ and σ^2. We see from (1.41) that the infinitesimal operator A is given by

$$A = \frac{1}{2} \frac{\partial^2}{\partial x^2}.$$

Putting

$$u(t, x) = \int u(y) f(t, x, y) \, dy$$

the Kolmogorov backward equation (1.45) reads

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} \frac{\partial^2 u(t, x)}{\partial x^2}.$$

This type of equation is called heat (or pure diffusion, which means without drift) equation in physics. The Fokker–Planck equation has the same form

$$\frac{\partial f(t, x, y)}{\partial t} = \frac{1}{2} \frac{\partial^2 f(t, x, y)}{\partial y^2}.$$

Of course, the above differential equation could also have been directly obtained using the fact that the transition density is, in view of $X_{t,x} = x + W_t$, given by

$$f(t, x, y) = \varphi_{0,t}(y - x) = \frac{1}{\sqrt{2\pi t}} \exp\left\{ -\frac{(y - x)^2}{2t} \right\}.$$

Example 1.3 *The Ornstein–Uhlenbeck process is defined to be a solution of the following stochastic differential equation*

$$dX_t = \mu X_t \, dt + \sigma \, dW_t.$$

To solve this equation we apply the Ito formula to the process $X_t \exp\{\mu t\}$ where we choose $u(t, x) = x \exp\{-\mu t\}$. The formula for $du(t, X_t)$ in Theorem 1.11 (Ito formula) gives

$$\begin{aligned} d(X_t \exp\{-\mu t\}) &= \frac{\partial u(t, X_t)}{\partial t} \, dt + \frac{\partial u(t, X_t)}{\partial x} \, dX_t + \frac{1}{2} \frac{\partial^2 u(t, X_t)}{\partial x^2} (dX_t)^2 \\ &= -\mu X_t \exp\{-\mu t\} \, dt + \exp\{-\mu t\} \, dX_t \\ &= \exp\{-\mu t\} \sigma \, dW_t. \end{aligned}$$

Hence

$$\exp\{-\mu t\}X_t - X_0 = \sigma \int_0^t \exp\{-\mu s\} \, dW_s$$

$$X_t = X_0 \exp\{\mu t\} + \sigma \int_0^t \exp\{\mu(t-s)\} \, dW_s.$$

The infinitesimal operator reads

$$A = \mu x \frac{\partial}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2}.$$

From the definition of the Ito integral one easily concludes that for any nonrandom function h the random variable

$$\int_0^t h(s) \, dW_s$$

has a normal distribution with expectation zero and variance $\int_0^t h^2(s) \, ds$. This means that the distribution of $\sigma \int_0^t \exp\{\mu(t-s)\} \, dW_s$ is a normal distribution with expectation zero and a variance given by

$$\sigma^2 \exp\{2\mu t\} \int_0^t \exp\{-2\mu s\} \, ds = -\frac{\sigma^2}{2\mu} \exp\{2\mu t\} \left[\exp\{-2\mu t\} - 1\right]$$

$$= -\frac{\sigma^2}{2\mu} \left[1 - \exp\{2\mu t\}\right] \longrightarrow -\frac{\sigma^2}{2\mu} \qquad \text{for } t \to \infty$$

if $\mu < 0$. In this case $X_0 \exp\{\mu t\}$ tends to zero. Hence for $\mu < 0$ the one-dimensional marginal distribution of X_t tends to a normal distribution with expectation zero and variance $-(\sigma^2/2\mu)$. One can show that this distribution, when used as an initial distribution of X_0, turns the Ornstein–Uhlenbeck process into a stationary process.

Example 1.4 *We consider the geometric Brownian motion that is defined by*

$$Y_t = \exp\{\mu t + \sigma W_t\}.$$

Put $X_t = \mu t + \sigma W_t$. We use the Ito formula in Theorem 1.11 with $u(x) = \exp\{x\}$. Hence by (1.32) and (1.33)

$$dY_t = \frac{\partial u(t, X_t)}{\partial t} \, dt + \frac{\partial u(t, X_t)}{\partial x} \, dX_t + \frac{1}{2} \frac{\partial^2 u(t, X_t)}{\partial x^2} (dX_t)^2$$

$$= u(X_t) \, dX_t + \frac{1}{2} u(X_t)(\mu \, dt + \sigma \, dW_t)^2$$

$$= Y_t \, dX_t + \frac{\sigma^2}{2} Y_t \, dt = Y_t \left(\mu + \frac{\sigma^2}{2}\right) dt + \sigma Y_t \, dW_t.$$

In particular, for $\mu = -\sigma^2/2$ *we get*

$$\mathrm{d}Y_t = \sigma Y_t \, \mathrm{d}W_t.$$

Hence we see from (1.23) that Y_t *has the constant expectation* $\mathbb{E}Y_0 = 1$.

These special cases of diffusion processes considered in the last three examples will be discussed in more detail in Chapter 6 (Wiener process or Brownian motion from Example 1.2), in Chapter 8 (Ornstein–Uhlenbeck process from Example 1.3) and in Chapter 11 as well as in Section 5.9 (geometric Brownian motion from Example 1.4).

1.6
Exercises

E 1.1 *Ito diffusion*
Write a computer program using the Euler discretization algorithm of the Ito stochastic differential equation (1.38) to study special cases of Ito diffusion such as the Wiener process, Brownian motion with constant drift, and especially geometric Brownian motion (see Examples 1.2–1.4 in Section 1.5). Start with a simulation of the Wiener process $\mathrm{d}X_t = \mathrm{d}W_t$ *using a discrete time interval* Δt *and normally distributed random numbers* $Z \sim \mathcal{N}(0, 1)$ *generated by the Box–Muller and/or the polar method. Check the known properties of the Wiener process by considering the Wiener difference* $\Delta W_t = W_{t+\Delta t} - W_t$ *over time step* $\Delta t = t + \Delta t - t$ *in the limit* $\Delta t \to 0$.

E 1.2 *Brownian paths in higher dimensions*
Study Brownian paths (or Wiener trails) in higher dimensions $\mathbb{R}^n (n \geq 2)$ *and show that the n-dimensional Brownian motion is isotropic by doing simulations of Brownian paths in* $\mathbb{R}^2$.

E 1.3 *Hausdorff dimension*
The Hausdorff dimension and the box-counting dimension of a Brownian trail in $\mathbb{R}^n (n \geq 2)$ *is equal to 2. Try to find the Hausdorff and box dimension for a graph (realization) of Brownian motion in* $\mathbb{R}^1$ *(one-dimensional case).*

E 1.4 *Stochastic process with constant drift and diffusion*
Find the Fokker–Planck equation for the stochastic process that satisfies the stochastic differential equation $\mathrm{d}X_t = -a \, \mathrm{d}t + b \, \mathrm{d}W_t$, *where a and b are constants and* $\mathrm{d}W_t = W_{t+dt} - W_t$ *is the increment of a Wiener process (also called white noise).*

E 1.5 *Stochastic Ornstein–Uhlenbeck process*
Consider Example 1.3 in Section 1.5 (Ornstein–Uhlenbeck process) in more detail and find the solution of the corresponding Fokker–Planck equation related to $\mathrm{d}u_t = -\mu \, u_t \, \mathrm{d}t + \sigma \, \mathrm{d}W_t$ *with non-negative constants* μ, σ *and given the initial condition* $u_{t=0} = u_0$. *Show that the probability density* $p(u, t)$ *becomes stationary and the so-called fluctuation–dissipation relation holds.*