$\boxed{\text{Stochastic Differential Equations}}$

(Stieltjes cursus, Leiden, 19-23 februari 2001)

ONNO VAN GAANS

JAN VAN NEERVEN

Thomas Stieltjes Institute for Mathematics

Chapter 1

The notion of noise

1.1 Introduction

Model problem

Consider a small particle in a fluid. The particle is subject to bombardments by the molecules of the fluid. On average the effects will cancel, but if the particle is very small, then there are differences on an observable time scale. The result is a very erratic movement of the particle.
For a model, let $U(t)$ $(\in \mathbb{R}^3)$ be the velocity of the particle at time t. Changes in the velocity are caused by friction and by the colliding fluid molecules which we assume to be a very random influence for the time being called 'noise':

$$U'(t) = -aU(t) + \text{'noise'}, \quad t \geqslant 0. \tag{1.1}$$

Here a is a positive constant. Let u_0 be the initial value:

$$U(0) = u_0.$$

Equation (1.1) is sometimes called the *Langevin equation*.
Our first objective is to introduce a mathematical model for the 'noise' and to formulate (1.1) rigorously. After that we will discuss the solution of the equation and further properties.

What is 'noise'?

Assume for simplicity that $a = 0$, $u_0 = 0$, and consider the problem in one dimension:

$$\begin{aligned} U'(t) &= \text{'noise'}, \quad t \geqslant 0, \\ U(0) &= 0. \end{aligned} \tag{1.2}$$

Intuitively, the 'noise' is a completely random influence as if at every moment a coin is tossed to decide in which direction the influence will be. To make this idea manageable, we will first consider discretizations.
Let $T > 0$ and let $n \in \mathbb{N}$. Divide the interval $[0, T]$ in n equal parts and let $h = T/n$. We toss n times with a coin and denote the outcomes by $\xi_1, \ldots, \xi_n \in \{-1, 1\}$ where 1 stands for 'Head' and -1 for 'Tail'. Now we set

$$\frac{U^{(n)}(kh) - U^{(n)}((k-1)h)}{h} = \alpha^{(n)}\xi_k, \quad k = 1, \ldots, n,$$

where $\alpha^{(n)}$ is some real scaling constant to be determined later on. Rewriting yields

$$\begin{aligned} U^{(n)}(kh) &= U^{(n)}((k-1)h) + h\alpha^{(n)}\xi_k \\ &= u_0 + \sum_{j=1}^{k} h\alpha^{(n)}\xi_j \\ &= \sum_{j=1}^{k} h\alpha^{(n)}\xi_j, \quad k = 1, \ldots, n. \end{aligned}$$

If we take $U^{(n)}$ on each interval $((k-1)h, kh)$ equal to its value at the right end point, then we obtain

$$\begin{aligned} U^{(n)}(t) &= U(0) + \sum_{j=1}^{\lceil t/h \rceil} h\alpha^{(n)}\xi_j \\ &= \sum_{j=1}^{\lceil nt/T \rceil} \tfrac{T}{n}\alpha^{(n)}\xi_j, \quad 0 < t \leqslant n. \end{aligned}$$

Here $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. Now we want to let $n \to \infty$ and obtain a suitable limit to define U.

To complete the above approach, we need a proper description of the outcomes of a sequence of coin tossings and a suitable notion of limit of sums of such outcomes. In the next section we introduce the notions of probability theory that deal with the above issues. They are not only needed to formulate the 'noise', but also to analyse the ensuing differential equation.

1.2 Probability theory

Probability space

Let S be a nonempty set. A *σ-algebra* (*σ-field*) in S is a collection $\mathscr{F}$ of subsets of S such that

(i) $\varnothing \in \mathscr{F}$,

(ii) $A \in \mathscr{F} \implies A^c \in \mathscr{F}$, and

(iii) $A_1, A_2, \ldots \in \mathscr{F} \implies \cup_{k=1}^{\infty} A_k \in \mathscr{F}$.

If S is a topological space, for instance a metric space, then the *Borel σ-algebra* is the σ-algebra generated by the open sets, i.e. the smallest σ-algebra containing all open sets of S. The members of the Borel σ-algebra are called the *Borel sets* of S. The Borel σ-algebra of $\mathbb{R}$ is denoted by $\mathscr{B}$.

A *measure* on a σ-algebra $\mathscr{F}$ is a mapping $\mu\colon\ \mathscr{F} \to [0, \infty]$ that is 0 at $\varnothing$ and *σ-additive*:

$$\mu(\cup_{k=1}^{\infty} A_k) = \textstyle\sum_{k=1}^{\infty} \mu(A_k)$$

for every pairwise disjoint collection $\{A_1, A_2, \ldots\}$ in $\mathscr{F}$. If $\mu(S) = 1$, then μ is called a *probability measure* on $\mathscr{F}$, or, less precisely, on S. A measure on a Borel-σ-algebra is called a *Borel measure*.

Definition 1.1. A *probability space* consists of a triple $(\Omega, \mathscr{F}, \mathbb{P})$ where

(i) Ω is a non empty set, called the *sample space* and its elements are called *sample points*,

(ii) $\mathscr{F}$ is a σ-algebra of subsets of Ω; these subsets are called *events*,

(iii) $\mathbb{P}(.)$ is a *probability measure* or briefly *probability* on $\mathscr{F}$.

The notion of a probability space serves to model random phenomena. The set S is the set of all possible configurations that may occur, $\mathscr{F}$ the collection of all events of which you might want to know the probability, and $\mathbb{P}$ assigns to each event its probability, i.e., the probability that an outcome of the random phenomenon is a member of that event.

Example 1.2. *Tossing with a fair coin.* Consider the probabilistic phenomenon of tossing n times with a fair coin. We can model the space of all possible outcomes (sample points) as $\Omega = \{H, T\}^n$, i.e., the sequences of length n consisting of the letters H ('head') and T ('tail'). The events that we are interested in are typically subsets of sequences that contain seven H's, that contain the string (H, T, T) but not (T, H, T), etc. To cover everything we might want to consider, take for $\mathscr{F}$ the collection of all subsets of Ω. Finally we define $\mathbb{P}$. Since we are tossing with a fair coin, every sequence is equally likely to occur. Therefore we should set the probability of a singleton equal to $1/2^n$. Then $\mathbb{P}$ is completely determined by the requirement that it should be σ-additive: $\mathbb{P}(A) = \#A/2^n$, $A \in \mathscr{F}$.

In case S is finite, one usually takes for $\mathscr{F}$ the collection of all subsets and defines $\mathbb{P}$ on the singletons. The general definition is of more use in case of an uncountable sample space, as it may not be possible to define a suitable probability measure on all subsets of S. For an (artificial) example we can think of the position of the valve of the front wheel of your bike after a long bike ride. In that case a reasonable model may be to take $S = [0, 2\pi)$, $\mathscr{F}$ the Borel-σ-algebra of S, and $\mathbb{P} = 1/2\pi$ times Lebesgue-measure.

Random variables and distributions

Many questions concerning a random phenomenon involve only one numerical quantity X, for instance the number of heads that occur when tossing a coin n times. Typical questions are with what probability X does not exceed a certain value, lies in a certain interval, etc.

Definition 1.3. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A function $X\colon\ \Omega \to \mathbb{R}$ is called a *random variable* if X is *measurable*, that means: for every Borel set B in $\mathbb{R}$ the set $\{\omega \in \Omega\colon\ X(\omega) \in B\}$ is a member of $\mathscr{F}$. We will denote

$$\{X \in B\} := \{\omega \in \Omega\colon\ X(\omega) \in B\}$$

and, accordingly,

$$\{X \leqslant x\} := \{\omega \in \Omega\colon\ X(\omega) \leqslant x\}, \qquad x \in \mathbb{R}.$$

If $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space and $X : \Omega \to \mathbb{R}$ is a random variable on Ω, then we can assign to every Borel set B of $\mathbb{R}$ the number $\mathbb{P}(\{X \in B\})$. Thus, a probability measure on $\mathbb{R}$ is defined and it is called the *law* or *distribution* of X. It is a central notion in probability theory. We will need a similar notion for vectors of random variables as well. Note that a function $X : \Omega \to \mathbb{R}^d$ is measurable with respect to $\mathscr{F}$ and the Borel-σ-algebra $\mathscr{B}_d$ of $\mathbb{R}^d$ if and only if each component X_i, $i = 1, \ldots, d$, of X is a random variable. We call such a function X a *random vector*.

Definition 1.4. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $X\colon \Omega \to \mathbb{R}^d$ be a random vector. Then:

(i) The collection of subsets of Ω of the form $\{X \in B\}$, $B \in \mathscr{B}_d$, is a σ-algebra, denoted by $\mathscr{F}(X)$ or $\sigma(X)$, called the σ-algebra *generated by* X. Note that $\mathscr{F}(X)$ is a subcollection of $\mathscr{F}$, since X is Borel measurable.

(ii) The measure μ_X defined by

$$\mu_X(B) := \mathbb{P}(\{X \in B\}), \qquad B \in \mathscr{B}_d,$$

is called the *law* or *distribution* of the random vector X. If $X_1, \dots, X_d$ denote the components of X, then μ_X is also called the *joint distribution* of the random variables $X_1, \dots, X_d$.

It is in computations often easier to deal with the distribution of a random vector than with the underlying probability measure of the probability space, especially when the distribution has a density with respect to the Lebesgue measure. The following proposition helps to reduce to distributions.

Proposition 1.5. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $X : \Omega \to \mathbb{R}^d$ a random vector, and $f\colon \mathbb{R}^d \to \mathbb{R}$ a Borel measurable nonnegative function. Let μ_X denote the distribution of X. Then*

$$\int_\Omega f(X)d\mathbb{P} = \int_{\mathbb{R}^d} f(x_1 \dots, x_d)\, d\mu_X(x_1, \dots, x_d).$$

The same assertion holds true for a Borel measurable function f such that $\int_\Omega |f(X)|d\mathbb{P} < \infty$.

Proof. If f is the indicator function of a Borel set B, then the equality is precisely the definition of $\mu_X(B)$. For step functions f the equality then follows by linearity of integration. An arbitrary nonnegative Borel measurable function is the pointwise limit of an increasing sequence of step functions and the equality then follows with aid of the monotone convergence theorem. The final assertion is established by splitting f in its positive and negative part. $\square$

Example 1.6. *Distribution of coin tossing.* Let $\Omega = \{H, T\}$, $\mathscr{F}$ the collection of all four subsets of Ω, and $\mathbb{P}$ the measure on $\mathscr{F}$ defined by $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = 1/2$. Define $\xi\colon \Omega \to \mathbb{R}$ by $\xi(H) = 1$, $\xi(T) = -1$. Then the distribution of ξ is the measure $(\delta_1 + \delta_{-1})/2$, where δ_a denotes the Dirac measure of the point a in $\mathbb{R}$, i.e., for $A \subset \mathbb{R}$ we have $\delta_a(A) = 1$ if $a \in A$ and $\delta_a(A) = 0$ otherwise.

Expectation and variance

Let X be a random variable such that each X_i is nonnegative or $\int_\Omega |X|d\mathbb{P} < \infty$. Then the *expectation* or *mean* of X is given by $\mathbb{E}(X) = \int_\Omega X d\mathbb{P}$. In view of Proposition 1.5, we have

$$\mathbb{E}(X) = \int_{\mathbb{R}} x\, d\mu_X(x),$$

where μ_X is the distribution of X. Observe that the map $X \mapsto \mathbb{E}(X)$ is linear.
If $\mathbb{E}(|X|^2) < \infty$, then $\operatorname{var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2)$ is called the *variance* of X, also denoted by $\sigma^2(X)$. We have

$$\operatorname{var}(X) = \int_\Omega X^2 d\mathbb{P} - (\mathbb{E}(X))^2.$$

If $X_1, \ldots, X_d$ are random variables, then the *covariance matrix* of $X_1, \ldots, X_d$ is the matrix $\mathrm{cov}\,(X_1, \ldots, X_d)$ whose coefficients are defined by

$$\mathrm{cov}\,(X_1, \ldots, X_d)_{ij} = \mathbb{E}\Big((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\Big), \quad i, j = 1, \ldots, d,$$

provided that $\mathbb{E}(X_i^2) < \infty$, $i = 1, \ldots, d$.

Independence

Let us say some words about the stochastic intuition behind independence before defining it.
The probability of an event can be interpreted as the relative frequency of occurences in a (infinitely) large number of samples. Thinking of two events, A and B, and a large number of samples, we can consider the probabilistic relationship between A and B by comparing the relative frequency of occurences of B among all the samples and among only those where A occurs. One extreme case we can imagine is that B occurs then and only then when A occurs. An other extreme case is that B occurs with the same frequency among the samples where A occurs as among all samples. Whether A occurs or not seems then of no importance to B. In the latter case, A and B are called *independent*. Since the fraction of the samples where both A and B occur equals the fraction of occurence of A times the fraction of occurence of B among these where A occurs, we see that for independent A and B we have: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Thus we arrive at the common definition of independency.

Definition 1.7. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. The σ-algebras $\mathscr{F}_1, \ldots, \mathscr{F}_n$ contained in $\mathscr{F}$ are said to be *independent* if for any choice of sets $A_1 \in \mathscr{F}_1, \ldots, A_n \in \mathscr{F}_n$ one has that

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \Pi_{k=1}^n \mathbb{P}(A_k).$$

Random variables $X_1, \ldots, X_n$ on $(\Omega, \mathscr{F}, \mathbb{P})$ are said to be *independent* if the σ-algebras generated by these random variables, $\mathscr{F}(X_1), \ldots, \mathscr{F}(X_n)$, are independent.
Clearly, $X_1, \ldots, X_n$ are independent if and only if

$$\mathbb{P}(\{X_1 \in B_1, \ldots, X_n \in B_n\}) = \Pi_{k=1}^n \mathbb{P}(\{X_k \in B_k\}) \tag{1.3}$$

for any sets $B_1, \ldots, B_n \in \mathscr{B}$.
A sequence $X_1, X_2, \ldots$ of random variables is called independent if for every $n \geqslant 1$, the random variables $X_1, \ldots, X_n$ are independent.

Random variables are independent if and only if their joint distribution is a product measure. The precise statement is in the proposition below. Recall that, given finite Borel measures $\mu_1, \ldots, \mu_n$ on $\mathbb{R}$, the *product measure* $\mu_1 \otimes \cdots \otimes \mu_n$ is the unique Borel measure on $\mathbb{R}^n$ that assigns to each 'hypercube' $B_1 \times \cdots \times B_n$ measure $\mu_1(B_1) \cdots \mu_n(B_n)$:

$$(\mu_1 \otimes \cdots \otimes \mu_n)(B_1 \times \cdots \times B_n) = \mu_1(B_1) \cdots \mu_n(B_n)$$

for every Borel sets $B_1, \ldots, B_n$ in $\mathbb{R}$.

Proposition 1.8. *Let $X_1, \ldots, X_n$ be random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mu_1, \ldots, \mu_n$ be their distributions. Then $X_1, \ldots, X_n$ are independent if and only if their joint distribution is the product measure $\mu_1 \otimes \cdots \otimes \mu_n$, i.e.,*

$$\mu_{(X_1, \ldots, X_n)} = \mu_1 \otimes \cdots \otimes \mu_n,$$

where $\mu_{(X_1, \ldots, X_n)}$ denotes the distribution of the random vector $(X_1, \ldots, X_n)$.

Proof. If $\mu_{(X_1,\ldots,X_n)} = \mu_1 \otimes \cdots \otimes \mu_n$, then for every Borel sets $B_1, \ldots B_n$ in $\mathbb{R}$ one has that

$$\begin{aligned}
\mathbb{P}(\{X_1 \in B_1, \ldots, X_n \in B_n\}) &= \mu_{(X_1,\ldots,X_n)}(B_1 \times \cdots \times B_n) \\
&= \mu_1(B_1) \cdots \mu_n(B_n) \\
&= \Pi_{k=1}^n \mathbb{P}(\{X_k \in B_k\}),
\end{aligned}$$

which means that $X_1, \ldots, X_n$ are independent.
Conversely, if $X_1, \ldots, X_n$ are independent, then for every Borel sets $B_1, \ldots, B_n$ in $\mathbb{R}$ one has that

$$\begin{aligned}
\mu_{(X_1,\ldots,X_n)}(B_1 \times \cdots \times B_n) &= \mathbb{P}(\{X_1 \in B_1, \ldots, X_n \in B_n\}) \\
&= \Pi_{k=1}^n \mathbb{P}(\{X_k \in B_k\}) \\
&= \mu_1(B_1) \cdots \mu_n(B_n),
\end{aligned}$$

so that $\mu_{(X_1,\ldots,X_n)}$ has to be the product measure $\mu_1 \otimes \cdots \otimes \mu_n$. □

Proposition 1.9. *Let X, Y be random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with $\mathbb{E}|X| < \infty$ and $\mathbb{E}|Y| < \infty$. If X and Y are independent, then $\mathbb{E}|XY| < \infty$ and*

$$\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y.$$

Proof. Let $\mu_{(X,Y)}$, μ_X, and μ_Y denote the distributions of (X, Y), X, and Y, respectively. As X and Y are independent we have that $\mu_{(X,Y)} = \mu_X \otimes \mu_Y$. First, Fubini for nonnegative functions yields

$$\begin{aligned}
\mathbb{E}|XY| &= \int_\Omega |X(\omega)|\,|Y(\omega)|\,d\mathbb{P}(\omega) = \int_{\mathbb{R}^2} |x|\,|y|\,d\mu_{(X,Y)}(x,y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} |x|\,|y|\,d\mu_X(x) d\mu_Y(y) \\
&= \int_{\mathbb{R}} |x|\,d\mu_X(x) \int_{\mathbb{R}} |y|\,d\mu_Y(y) = \mathbb{E}X\mathbb{E}Y,
\end{aligned}$$

which is a finite number. Hence the function $(x, y) \mapsto xy$ is integrable with respect to $\mu_X \otimes \mu_Y$. Now Fubini for integrable functions yields

$$\begin{aligned}
\mathbb{E}XY &= \int_\Omega X(\omega)Y(\omega)\,d\mathbb{P}(\omega) = \int_{\mathbb{R}^2} xy\,d\mu_{(X,Y)}(x,y) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} xy\,d\mu_X(x) d\mu_Y(y) \\
&= \int_{\mathbb{R}} x\,d\mu_X(x) \int_{\mathbb{R}} y\,d\mu_Y(y) = \mathbb{E}X\mathbb{E}Y.
\end{aligned}$$

□

By means of product measures we can construct independent random variables. Indeed, let X_1 and X_2 be two random variables, each of them living on its own probability space, say $(\Omega_1, \mathscr{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathscr{F}_2, \mathbb{P}_2)$. Let $\Omega := \Omega_1 \times \Omega_2$, let $\mathscr{F}$ be the σ-algebra in Ω generated by the sets $A_1 \times A_2$, with $A_1 \in \mathscr{F}_1$, $A_2 \in \mathscr{F}_2$, and let $\mathbb{P} := \mathbb{P}_1 \otimes \mathbb{P}_2$. Define $\bar{X}_1(\omega_1, \omega_2) := X_1(\omega_1)$ and $\bar{X}_2(\omega_1, \omega_2) := X_2(\omega_2)$. Then $\bar{X}_1$ and $\bar{X}_2$ are independent random variables on Ω and their distributions equal the distributions of X_1 and X_2, respectively.

It is clear how the construction can be extended to any finite number of random variables. To construct a sequence of independent random variables, let $X_1, X_2, \dots$ be a sequence of random variables with each X_k defined on a probability space $(\Omega_k, \mathscr{F}_k, \mathbb{P}_k)$, $k = 1, 2, \dots$. Take $\Omega := \Pi_{k=1}^{\infty} \Omega_k = \Omega_1 \times \Omega_2 \times \cdots$, the cartesian product. A subset of Ω is called a *(rectangular) cylinder set* if it is of the form $A \times \Omega_{k+1} \times \Omega_{k+2} \times \cdots$, where $A = A_1 \times \cdots \times A_k$ with $A_1 \in \mathscr{F}_1 \dots, A_k \in \mathscr{F}_k$ and $k \geqslant 1$. Let $\mathscr{F}$ be the σ-algebra in Ω generated by the cylinder sets. Then it can be shown (see, e.g., Halmos[8, §38 Thm B, p.157]) that there is one and only one probability measure $\mathbb{P}$ on $\mathscr{F}$ such that

$$\mathbb{P}(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \cdots) = \mathbb{P}_1(A_1) \cdots \mathbb{P}_n(A_n)$$

for every $n \geqslant 1$ and $A_1 \in \mathscr{F}_1, \dots, A_n \in \mathscr{F}_n$. The space $(\Omega, \mathscr{F}, \mathbb{P})$ is called the *product* of the probability spaces $(\Omega_k, \mathscr{F}_k, \mathbb{P}_k)$, $k = 1, 2, \dots$. By defining $\bar{X}_k(\omega_1, \omega_2, \dots) := X_k(\omega_k)$, we find a sequence of independent random variables with the same distributions as the X_k, $k = 1, 2, \dots$.

An interesting special case is that we have a sequence of copies of the same random variable. We then obtain a sequence of independent identically distributed random variables, usually refered to by the term *i.i.d. sequence*.

We have now the tools available to construct the sequence of coin tossing $\xi_1, \xi_2, \dots$ as mentioned in Section 1.1. Indeed, take for each $k \in \mathbb{N}$ the sample space $\Omega_k = \{H, T\}$, $\mathscr{F}$ the collection of all four subsets of Ω, and $\mathbb{P}$ the probability measure on $\mathscr{F}$ defined by $\mathbb{P}_k(\{H\}) = \mathbb{P}_k(\{T\}) = 1/2$. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be the product space of the $(\Omega_k, \mathscr{F}_k, \mathbb{P}_k)$, $k = 1, 2, \dots$. Define the random variables $\xi_1, \xi_2, \dots$ on $(\Omega, \mathscr{F}, \mathbb{P})$ by

$$\xi_k(\omega_1, \omega_2, \dots) = \begin{cases} 1 & \text{if} \quad \omega_k = H \\ -1 & \text{if} \quad \omega_k = T, \end{cases} \qquad (\omega_1, \omega_2, \dots) \in \Omega,$$

for $k = 1, 2, \dots$.

In the remainder of this section we will present some computational tools, Gaussian random variables —which will occur in the limit of sums of the ξ_k— and the notion of convergence in distribution.

Fourier transforms of distributions

A useful tool for manipulations with distributions is the Fourier transform.

Definition 1.10. Given a Borel probability measure μ on $\mathbb{R}^d$, its *Fourier transform* $\hat{\mu}$ is the function from $\mathbb{R}^d$ to $\mathbb{C}$ defined by

$$\hat{\mu}(u_1, \dots, u_m) := \int_{\mathbb{R}^d} e^{-i\langle u, x\rangle} d\mu(x), \quad u_1, \dots, u_m \in \mathbb{R}, \tag{1.4}$$

where $\langle u, x\rangle = u_1 x_1 + \cdots + u_d x_d$.

Probabilists usually consider the function $u \mapsto \int_{\mathbb{R}^d} e^{i\langle u, x\rangle} d\mu(x)$ and call it the *characteristic function* of μ.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $X = (X_1, \dots, X_d)$ be a random vector. If μ_X denotes the distribution of X, we have in view of Proposition 1.5 that

$$\begin{aligned} \hat{\mu}_X(u_1, \dots, u_d) &= \int_{\mathbb{R}^d} e^{-i\langle u,x\rangle} d\mu_X(x) \\ &= \mathbb{E}(e^{-i\langle u,X\rangle}), \quad u_1, \dots, u_d \in \mathbb{R}. \end{aligned}$$

Proposition 1.11. *Let μ and ν be Borel probability measures on $\mathbb{R}^d$ and $\hat{\mu}$ and $\hat{\nu}$ their Fourier transforms. If $\hat{\mu} = \hat{\nu}$, then $\mu = \nu$.*

Proposition 1.12. *Let X and Y be random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\mu_{(X,Y)}$, μ_X, and μ_Y denote the distribution of (X, Y), X, and Y, respectively. If X and Y are independent, then*

$$\hat{\mu}_{(X,Y)}(u, v) = \hat{\mu}_X(u)\hat{\mu}_Y(v) \quad \text{for all } u, v \in \mathbb{R}.$$

Proof.

$$\begin{aligned} \hat{\mu}_{(X,Y)}(u, v) &= \int_{\mathbb{R}^2} e^{-i(ux+vy)} d\mu_{(X,Y)}(x, y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-iux} e^{-ivy} d\mu_X(x) d\mu_Y(y) \\ &= \hat{\mu}_X(u)\hat{\mu}_Y(v). \end{aligned}$$

□

Gaussian measures

Definition 1.13. (i) A *Gaussian measure μ on $\mathbb{R}^d$* is a Borel probability measure on $\mathbb{R}^d$ with Fourier transform $\hat{\mu}\colon \mathbb{R}^d \to \mathbb{C}$ of the form

$$\hat{\mu}(u_1, \dots, u_d) = e^{-i\langle u,m\rangle - \frac{1}{2}\langle Cu,u\rangle},$$

where $m = (m_1, \dots, m_d) \in \mathbb{R}^d$ and C is a symmetric nonnegative definite $d \times d$ matrix. The measure μ is denoted by $N(m, C)(dx)$.

(ii) A *random vector* $(X_1, \dots, X_d)$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called *Gaussian* (or *normally distributed*) if its distribution is a Gaussian measure on $\mathbb{R}^d$.

(iii) A collection of random variables $\{X_\alpha\}_{\alpha \in \mathscr{I}}$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, indexed by a nonempty index set $\mathscr{I}$, is called a *Gaussian system* (or family) if the random vector $(X_{\alpha_1}, \dots, X_{\alpha_n})$ is Gaussian for every $n \geqslant 1$ and all indices $\alpha_1, \dots, \alpha_n$ chosen from $\mathscr{I}$.

Proposition 1.14. *If $(X_1, \dots, X_d)$ is a Gaussian vector with distribution μ such that*

$$\hat{\mu}(u_1, \dots, u_d) = \exp(-i\langle u, m\rangle - \frac{1}{2}\langle u, Cu\rangle), \qquad u \in \mathbb{R}^d,$$

then

$$\begin{aligned} m_k &= \int_\Omega X_k d\mathbb{P}, \quad k = 1, \dots, d, \text{ and} \\ C_{kl} &= \int_\Omega (X_k - m_k)(X_l - m_l) d\mathbb{P}, \quad 1 \leqslant k, l \leqslant d. \end{aligned}$$

Hence $m = (m_1, \ldots, m_d)$ *is the mean value of* $(X_1, \ldots, X_d)$ *and* C *is the covariance matrix of* $(X_1, \ldots, X_d)$*. If* $m = 0$*, then the random vector is called* centered.

Proof. By definition,

$$\hat{\mu}(u) = \int_{\mathbb{R}^d} e^{-i\langle u,x\rangle} d\mu(x), \quad u = (u_1, \ldots, u_n) \in \mathbb{R}^d.$$

With aid of Lebesgue's dominated convergence theorem it follows that

$$\begin{aligned} \frac{\partial}{\partial u_k}\hat{\mu}(u) &= \int_{\mathbb{R}^d} \frac{\partial}{\partial u_k} e^{-i\langle u,x\rangle} d\mu(x) \\ &= \int_{\mathbb{R}^d} -ix_k e^{-i\langle u,x\rangle} d\mu(x). \end{aligned}$$

For $u = 0$ we find

$$\frac{\partial}{\partial u_k}\hat{\mu}(0) = -i \int_{\mathbb{R}^d} x_k d\mu(x) = -i\mathbb{E}X_k.$$

On the other hand, the given expression for $\hat{\mu}$ yields

$$\frac{\partial}{\partial u_k}\hat{\mu}(0) = -im_k.$$

Hence $\mathbb{E}X_k = m_k$.
Similarly,

$$\frac{\partial^2}{\partial u_k \partial u_l}\hat{\mu}(0) = \int_{\mathbb{R}^d} -x_k x_l e^{-i\langle u,x\rangle} d\mu(x) = -\mathbb{E}X_k X_l$$

and

$$\frac{\partial^2}{\partial u_k \partial u_l}\hat{\mu}(0) = -m_k m_l - C_{kl},$$

hence

$$\mathbb{E}(X_k - m_k)(X_l - m_l) = C_{kl}.$$

□

Sometimes the following explicit formula for the density of the distribution of a Gaussian measure with respect to the Lebesgue measure is useful.

Proposition 1.15. *Let* μ *be a Gaussian measure on* $\mathbb{R}^d$ *with mean value* $m \in \mathbb{R}^d$ *and covariance (matrix)* C*. Then* μ *is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^d$ *if and only if* C *is nonsingular. In that case the density of* μ *is given by*

$$p(x) = \frac{1}{(2\pi)^{d/2}(\det C)^{1/2}} \exp\{-\tfrac{1}{2}\langle x - m, C^{-1}(x - m)\rangle\}, \quad x \in \mathbb{R}^d.$$

(See Shiryayev[11, §II.13.2, p. 296–298].)

Exercise 1.16. Verify that

$$\int_{\mathbb{R}} e^{-iux} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}} dx = e^{-ium - \frac{1}{2}\sigma^2 u^2}, \quad u \in \mathbb{R}.$$

Exercise 1.17. Let X be a Gaussian random variable with mean 0 and variance 1. Let $p \geqslant 0$.

1. Show that
$$\mathbb{E}|X|^p = \tfrac{1}{\sqrt{\pi}} 2^{\frac{p}{2}} \Gamma(\tfrac{p+1}{2}),$$
where Γ denotes the gamma-function: $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$, $z > 0$.

2. Show that
$$\mathbb{E}X^{2n} = (2n-1)(2n-3)\cdots 1, \quad n = 1, 2, \ldots.$$

If $(X_1, \ldots, X_d)$ is a Gaussian random vector, then $\lambda_1 X_1 + \cdots + \lambda_d X_d$ is a Gaussian random variable. In particular with $\lambda_i = \delta_{ij}$, X_i is a Gaussian random variable for all $i = 1, \ldots, d$. More generally we have the following.

Proposition 1.18. *Let $X = (X_1, \ldots, X_n)$ be a Gaussian random vector with expectation $m \in \mathbb{R}^d$ and covariance C. Let $\phi\colon\ \mathbb{R}^d \to \mathbb{R}^n$ be defined by $\phi(x) = a + Ax$ where $a \in \mathbb{R}^n$ and A is an $n \times d$ matrix. Then $\phi(X)$ is a Gaussian random vector with mean value $a + Am$ and covariance matrix ACA^T.*

Proof. Let μ be the distribution of X and let $u \in \mathbb{R}^n$. Then
$$\begin{aligned}
\mathbb{E}e^{-i\langle u, \phi(X)\rangle} &= \int_{\mathbb{R}^d} e^{-i\langle u, \phi(x)\rangle} d\mu(x) \\
&= \int_{\mathbb{R}^d} e^{-i\langle u, a\rangle} e^{-i\langle u, Ax\rangle} d\mu(x) \\
&= e^{-i\langle u, a\rangle} \int_{\mathbb{R}^d} e^{-i\langle A^T u, x\rangle} d\mu(x) \\
&= e^{-i\langle u, a\rangle} e^{-i\langle A^T u, m\rangle} e^{-\frac{1}{2}\langle A^T u, CA^T u\rangle} \\
&= e^{-i\langle a + Am, u\rangle - \frac{1}{2}\langle ACA^T u, u\rangle}.
\end{aligned}$$
□

If $X_1, \ldots, X_d$ are Gaussian random variables, then $\lambda_1 X_1 + \cdots + \lambda_d X_d$ may not be Gaussian.

Example 1.19. If X_1, X_2 are independent $N(0,1)$ and
$$(X, Y) := \begin{cases} (X_1, |X_2|) & \text{if } X_1 \geqslant 0, \\ (X_1, -|X_2|) & \text{if } X_1 < 0, \end{cases}$$
then X and Y are both Gaussian but (X, Y) is not.

Proposition 1.20. *(i) Let $(X_1, \ldots, X_d)$ be a random vector. Then $(X_1, \ldots, X_d)$ is a Gaussian random vector if and only if $\lambda_1 X_1 + \cdots + \lambda_d X_d$ is a Gaussian random variable for every $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$.*

(ii) Let $(X_1, \ldots, X_d)$ be Gaussian. Then $X_1, \ldots, X_d$ are independent if and only if they are uncorrelated*, i.e., the covariance matrix*
$$\Big(\mathbb{E}((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j))\Big)_{i,j}$$
is a diagonal matrix.

(iii) If $X_1, \ldots, X_d$ are independent Gaussian random variables, then $(X_1, \ldots, X_d)$ is Gaussian.
(See Shiryayev[11, Thm 1, p.299].)

Convergence in distribution

Now we turn to convergence of random vectors or rather convergence of their distributions. The notion of convergence that turns out to be most useful is based on 'weak convergence' of measures. Let $C_b(\mathbb{R}^d)$ denote the space of bounded, continuous functions from $\mathbb{R}^d$ to $\mathbb{C}$. Let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on $\mathbb{R}^d$. The sequence $(\mu_n)_n$ is said to *converge weakly* to μ if

$$\int_{\mathbb{R}^d} f \, d\mu_n \to \int_{\mathbb{R}^d} f \, d\mu \text{ as } n \to \infty \text{ for all } f \in C_b(\mathbb{R}^d).$$

Weak convergence is equivalent to pointwise convergence of the Fourier transforms:

Proposition 1.21. *Let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on $\mathbb{R}^d$. The following two statements are equivalent:*

(i) $\int_{\mathbb{R}^d} f \, d\mu_n \to \int_{\mathbb{R}^d} f \, d\mu$, *for all* $f \in C_b(\mathbb{R}^d)$,

(ii) $\int_{\mathbb{R}^d} e^{-i\langle u,x\rangle} d\mu_n(x) \to \int_{\mathbb{R}^d} e^{-i\langle u,x\rangle} d\mu(x)$, *for all* $u_1, \ldots, u_d \in \mathbb{R}$, *i.e.,* $\hat{\mu}_n$ *converges pointwise to* $\hat{\mu}$ *on* $\mathbb{R}^d$.

(See Stromberg[12, Cor. 2.8, p. 25].)

A sequence of random vectors is said to converge in distribution if their distributions converge weakly. There is a more direct way to state this. For a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a random vector X on Ω taking values in $\mathbb{R}^d$ with distribution μ_X, we have that

$$\mathbb{E}(f(X)) = \int_{\Omega} f(X) d\mathbb{P} = \int_{\mathbb{R}^d} f(x_1, \ldots, x_n) d\mu_X(x), \text{ for all } f \in C_b(\mathbb{R}^d).$$

So we can rewrite the convergence in terms of random vectors in the following way.

Definition 1.22. Let $X, X^{(1)}, X^{(2)}, \ldots$ be random vectors taking values in $\mathbb{R}^d$. The sequence $\{X^{(n)}\}_{n\geqslant 1}$ *converges in distribution* (in law) to X, notation $X^{(n)} \Rightarrow X$ or $X^{(n)} \xrightarrow{\mathrm{D}} X$, if

$$\mathbb{E}(f(X^{(n)})) \to \mathbb{E}(f(X)), \text{ for every } f \in C_b(\mathbb{R}^d). \tag{1.5}$$

Remark 1.23. Proposition 1.21 implies *Levy's Theorem*: we have $X^{(n)} \xrightarrow{\mathrm{D}} X$ if and only if $\mathbb{E}(e^{-i\langle u,X^{(n)}\rangle}) \to \mathbb{E}(e^{-i\langle u,X\rangle})$ for all $u_1, \ldots, u_d \in \mathbb{R}$, where $X^{(n)}$ and X are random vectors taking values in $\mathbb{R}^d$. In words: convergence in distributions is equivalent to pointwise convergence of the Fourier transforms of the distributions.

Proposition 1.24. *Let $\{\mu_n\}_{n\geqslant 1}$ be a sequence of Gaussian measures on $\mathbb{R}^d$ with expectations $\{m_n\}_n$ and covariances $\{C_n\}_n$. Then $\mu_n \xrightarrow{\mathrm{D}} \mu$ for some Borel probability measure μ on $\mathbb{R}^d$ if and only if $m_n \to m$ and $C_n \to C$ for some $m \in \mathbb{R}^d$ and some $d \times d$ matrix C. If this is the case, then μ is Gaussian with expectation m and covariance C.*
(See Shiryayev[11, §II.13.5, p. 302–303].)

Central Limit Theorem

The importance of Gaussian measures is that they appear as the result of an addition of many independent random influences.

Theorem 1.25. *Let* $\{(X_1^{(n)}, X_2^{(n)}, \ldots, X_d^{(n)})\}_{n\geqslant 1}$ *be a sequence of independent random vectors, all having the same distribution with zero mean and covariance matrix* C*. Then*

$$\tfrac{1}{\sqrt{n}}\left(\sum_{k=1}^{n} X_1^{(k)}, \ldots, \sum_{k=1}^{n} X_d^{(k)}\right) \xrightarrow{\mathrm{D}} (X_1, \ldots, X_d),$$

where $(X_1, \ldots, X_d)$ *is a Gaussian random vector with mean zero and covariance matrix* C*.*
(See Breiman[2, Thm 11.10, p. 238].)

1.3 A mathematical model for 'noise'

Let $\xi_1, \xi_2, \ldots$ be an i.i.d. sequence of random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with

$$\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = 1/2.$$

For each n we have as discretization of the function U of Section 1.1:

$$U^{(n)}(t) = \sum_{k=1}^{\lceil nt/T\rceil} \tfrac{T}{n}\alpha^{(n)}\xi_k, \quad 0 < t \leqslant T.$$

In order to apply the Central Limit Theorem, we have to take $\alpha^{(n)} = O(\sqrt{\lceil nt/T\rceil}) = O(\sqrt{n})$. To get a limit that is independent of T we choose

$$\alpha^{(n)} = \sqrt{\tfrac{n}{T}}.$$

Then the Central Limit Theorem yields that

$$U^{(n)}(t) = \sqrt{\tfrac{T}{n}} \sum_{k=1}^{\lceil nt/T\rceil} \xi_k = \sqrt{\tfrac{T}{n}\lceil nt/T\rceil}\,\frac{1}{\sqrt{\lceil nt/T\rceil}} \sum_{k=1}^{\lceil nt/T\rceil} \xi_k$$

converges to $\sqrt{t}$ times a Gaussian variable with mean 0 and variance equal to

$$\operatorname{var}\xi_1 = 1.$$

Hence the limit in distribution of $(U^{(n)}(t))_{t\geqslant 0}$ should be a family $(U(t))_{t\geqslant 0}$ such that for every t the random variable $U(t)$ is Gaussian with mean 0 and variance t.
We can also compute what the joint distributions should be. It turns out to be more informative to consider joint distributions of increments. Let $0 = t_0 < t_1 < \cdots < t_m$ and define its *increments*

$$\begin{aligned} V_k^{(n)} &:= U^{(n)}(t_k) - U^{(n)}(t_{k-1}), \quad k = 1, \ldots, m, \text{ and} \\ V^{(n)} &:= (V_1^{(n)}, \ldots, V_m^{(n)}). \end{aligned}$$

Note that

$$V_k^{(n)} = \sum_{j=\lceil nt_{k-1}/T\rceil+1}^{\lceil nt_k/T\rceil} \sqrt{\tfrac{T}{n}}\xi_j, \quad k = 1, \ldots, m.$$

Proposition 1.26. *The sequence $V^{(n)}$ as defined above converges in distribution to a random vector $(X_1, \dots, X_m)$ where X_k is a Gaussian with mean 0 and variance $t_k - t_{k-1}$, $k = 1, \dots, m$.*

Proof. Let $\mu_{V^{(n)}}$ denote the (m-dimensional) distribution of $V^{(n)}$ and $\mu_{V_k^{(n)}}$ the distribution of $V_k^{(n)}$, $k = 1, \dots, m$. Choose n so large that $\lceil nt_{k-1}/T \rceil < \lceil nt_k/T \rceil$ for $k = 1, \dots, d$. Then because the ξ_j are independent, the $V_k^{(n)}$ are independent and therefore

$$\hat{\mu}_{V^{(n)}}(u_1, \dots, u_m) = \Pi_{k=1}^m \hat{\mu}_{V_k^{(n)}}(u_k), \quad u_1, \dots, u_n \in \mathbb{R}.$$

Moreover, for each k and each $v \in \mathbb{R}$:

$$\hat{\mu}_{V_k^{(n)}}(v) = \Pi_{j=\lceil nt_{k-1}/T \rceil + 1}^{\lceil nt_k/T \rceil} \hat{\mu}_{\sqrt{T/n}\,\xi_j}(v) = \hat{\mu}_{\sqrt{T/n}\,\xi_1}(v)^{\lceil nt_k/T \rceil - \lceil nt_{k-1}/T \rceil}.$$

Note that

$$\begin{aligned} \hat{\mu}_{\sqrt{T/n}\,\xi_1}(v) &= \int_{\mathbb{R}} e^{-iv\sqrt{T/n}\,x}\, d\mu_{\xi_1}(x) \\ &= \int_{\mathbb{R}} e^{-iv\sqrt{T/n}\,x} (\tfrac{1}{2}\delta_{-1} + \tfrac{1}{2}\delta_1)(dx) \\ &= \cos(v\sqrt{T/n}). \end{aligned}$$

By straightforward calculation it can be shown that

$$[\cos(\beta/\sqrt{n})]^n \to e^{-\beta^2/2} \quad \text{as } n \to \infty.$$

Thus we have that

$$\begin{aligned} \hat{\mu}_{V_k^{(n)}}(u) &= \cos(u\sqrt{T}/n)^{\lceil nt_k/T \rceil - \lceil nt_{k-1}/T \rceil} \\ &\to e^{-\frac{1}{2}u^2 T(t_k - t_{k-1})/T}, \quad k = 1 \dots, m, \end{aligned}$$

and therefore

$$\hat{\mu}_{V^{(n)}}(u_1, \dots u_m) \to \Pi_{k=1}^m e^{\frac{1}{2}u_k^2(t_k - t_{k-1})} = e^{-\frac{1}{2}\sum_{k=1}^m u_k^2(t_k - t_{k-1})}.$$

Hence $\hat{\mu}_{V^{(n)}}$ converges pointwise to the Fourier transform of a Gaussian vector $(X_1, \dots, X_m)$ with means zero and a covariance matrix that is diagonal with on the diagonal entries the numbers $t_k - t_{k-1}$. This means that $X_1, \dots, X_m$ are independent Gaussians with

$$\mathbb{E}X_k = 0 \quad \text{and} \quad \operatorname{var} X_k = t_k - t_{k-1}, \quad k = 1, \dots, m.$$

□

It follows that the limit $U(t)$ of $U^{(n)}(t)$, $t \geqslant 0$, should be such that for any $0 = t_0 < t_1 < \cdots < t_m$ the increments $U(t_1) - U(t_0), \dots, U(t_m) - U(t_{m-1})$ are independent and each increment $U(t) - U(s)$ ($t \geqslant s$) is Gaussian with mean 0 and variance $t - s$. Does such a family of random variables exist? The answer is affirmative and the family is most famous: the Brownian motion.

Exercise 1.27. Use Proposition 1.20 to show that the family $(U(t))_{t \geqslant 0}$ should be a Gaussian system.

1.4 Brownian motion

A *stochastic process* is a family $(X_\lambda)_{\lambda\in\Lambda}$ of random variables all defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$. The index sets that we will encounter will almost always be intervals of $\mathbb{R}$, mainly $[0,\infty)$ and $[0,T]$ for some $T > 0$.
If $(Y_\lambda)_{\lambda\in\Lambda}$ is a family on $(\Omega, \mathscr{F}, \mathbb{P})$ and $Y_\lambda = X_\lambda$ $\mathbb{P}$-a.e. for every $\lambda \in \Lambda$, then $(Y_\lambda)_{\lambda\in\Lambda}$ is called a *version* of $(X_\lambda)_{\lambda\in\Lambda}$ (and $(X_\lambda)_{\lambda\in\Lambda}$ a version of $(Y_\lambda)_{\lambda\in\Lambda}$, of course).
For each $\omega \in \Omega$, the map $\lambda \mapsto X_\lambda(\omega)$ is called a *sample path* of the process $(X_\lambda)_{\lambda\in\Lambda}$. If Λ is an interval of $\mathbb{R}$, an important issue is whether $(X_\lambda)_{\lambda\in\Lambda}$ has continuous paths, or rather whether there exists a version of it with continuous paths. In our model, U is the velocity of a particle and its paths should be continuous in time.

Definition 1.28. A *normalized Brownian motion* (or *Wiener process*) in $\mathbb{R}$ is a family of random variables $\{W_t\}_{t\geqslant 0}$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that

(i) $W_0(\omega) = 0$ for almost all $\omega \in \Omega$,

(ii) for every $s, t \geqslant 0$, $W_{s+t} - W_s$ is a Gaussian random variable with mean 0 and variance t,

(iii) for every $0 = t_0 < t_1 < \cdots < t_n$ the random variables $W_{t_k} - W_{t_{k-1}}$, $k = 1, \dots, n$, are independent,

(iv) for almost every $\omega \in \Omega$ the *sample path* $t \mapsto W_t(\omega)$ from $[0,\infty)$ to $\mathbb{R}$ is continuous.

Because $W_0 = 0$, condition (iii) of the definition is equivalent to

(iii)' $W(t) - W(s)$ is independent of $W(r)$ for every $0 \leqslant r \leqslant s \leqslant t$.

Theorem 1.29. *There exists a Brownian motion in* $\mathbb{R}$.
(See Breiman[2].)

Remark 1.30. One can take for the underlying probability space of Brownian motion for Ω the set $C_0[0,\infty)$ of continuous functions on $[0,\infty)$ that vanish at 0, for $\mathscr{F}$ the Borel-σ-algebra with respect to the metric

$$d(\omega,\eta) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{d_k(\omega,\eta)}{1 + d_k(\omega,\eta)},$$

where

$$d_k(\omega,\eta) := \max_{0\leqslant t\leqslant k} |\omega(t) - \eta(t)|, \quad \omega, \eta \in \Omega,$$

and for the random variables

$$X(t)(\omega) = \omega(t), \quad t \in [0,\infty),\ \omega \in \Omega.$$

The measure $\mathbb{P}$ on Ω such that $(X(t))_{t\geqslant 0}$ is Brownian motion is called the *Wiener measure*.
Brownian motion is unique in the sense that if $(B(t))_{t\geqslant 0}$ is a Brownian motion in $\mathbb{R}$ with underlying probability space $(\Omega', \mathscr{F}', \mathbb{P}')$, then its *distribution*, which is the measure μ_B on $C_0[0,\infty)$ defined by

$$\begin{aligned} &\mu_B(\{f \in C_0[0,\infty):\ f(t_1) \in A_1, \dots, f(t_m) \in A_m\}) \\ &\quad = \mathbb{P}'(\{\omega \in \Omega':\ B(t_1)(\omega) \in A_1, \dots, B(t_m) \in A_m\}), \end{aligned}$$

where $0 < t_1 < \cdots < t_m$ and $A_1, \dots, A_m$ are Borel sets in $\mathbb{R}$, equals the Wiener measure.

Proposition 1.31. *Let $(W_t)_{t\geqslant 0}$ be normalized Brownian motion in $\mathbb{R}$. Then*

(i) $\mathbb{E}W_tW_s = s \wedge t$ *for every* $s,t \geqslant 0$.

(ii) If $0 = t_0 < t_1 < t_2 < \cdots < t_m$, *and* $B_1, \ldots, B_m$ *are Borel subsets of* $\mathbb{R}$, *then*

$$\mathbb{P}(\{W_{t_1} \in B_1, \ldots, W_{t_m} \in B_m\}) = \int_{B_1} \cdots \int_{B_m} \Pi_{k=1}^m p(t_k - t_{k-1}; x_{k-1}, x_k) dx_1 \cdots dx_m,$$

where

$$p(s; x, y) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2}\frac{(y-x)^2}{s}}, \quad s > 0,\ x, y \in \mathbb{R}.$$

This formula is called the Einstein-Smoluchowski formula.

Proof. (i): Assume that $t \geqslant s$. By independence of increments we find

$$\mathbb{E}W_tW_s = \mathbb{E}W_sW_s + \mathbb{E}W_s(W_t - W_s) = s + \mathbb{E}W_s\mathbb{E}(W_t - W_s) = s.$$

(ii): Let $V_1 := W_{t_1}$, $V_2 := W_{t_2} - W_{t_1}, \ldots$, $V_m := W_{t_m} - W_{t_{m-1}}$. For any Borel set A of $\mathbb{R}^m$ we have that

$$\begin{aligned}
&\mathbb{P}\{(W_{t_1}, W_{t_2}, \ldots, W_{t_m}) \in A\} \\
&= \mathbb{P}\{(V_1, V_1 + V_2, \ldots, V_1 + V_2 + \cdots + V_m) \in A\} \\
&= \int_{\mathbb{R}^m} \mathbf{1}_A(y_1, y_1 + y_2, \ldots, y_1 + \cdots + y_m)\, d\mu_{V_1}(y_1) \cdots d\mu_{V_m}(y_m) \\
&= \frac{1}{\sqrt{\Pi_{k=1}^m 2\pi(t_k - t_{k-1})}} \int_{\mathbb{R}^m} \mathbf{1}_A(y_1, y_1 + y_2, \ldots, y_1 + \cdots + y_m) \\
&\qquad\qquad \cdot e^{-\frac{1}{2}\sum_{k=1}^m \frac{y_k^2}{2(t_k - t_{k-1})}} dy_1 \cdots dy_m \\
&= \frac{1}{\sqrt{\Pi_{k=1}^m 2\pi(t_k - t_{k-1})}} \int_{\mathbb{R}^m} \mathbf{1}_A(x_1, x_2, \ldots, x_m) e^{-\frac{1}{2}\sum_{k=1}^m \frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}} dx_1 \cdots dx_m \\
&= \frac{1}{\sqrt{\Pi_{k=1}^m 2\pi(t_k - t_{k-1})}} \int_A e^{-\frac{1}{2}\sum_{k=1}^m \frac{(x_k - x_{k-1})^2}{2(t_k - t_{k-1})}} dx_1 \cdots dx_m.
\end{aligned}$$

In particular, if $A = B_1 \times B_2 \times \cdots \times B_m$, we find the formula we wanted to prove. □

Exercise 1.32. Show that the sequence $U^{(n)}$ as constructed in Section 1.1 converges in distribution to a Brownian motion in the sense that for every $0 < t_1 < \cdots < t_m \leqslant T$ the distributions of $(U^{(n)}(t_1), \ldots, U^{(n)}(t_m))$ converge to the distribution of $(W(t_1), \ldots, W(t_m))$, where $(W(t))_{t\geqslant 0}$ is a Brownian motion.

1.5 Setting

We return to equation (1.2) of Section 1.1. We have seen that for $\alpha^{(n)} = \sqrt{n/T}$ the sequence $U^{(n)}$ converges to Brownian motion in the sense of Exercise 1.32 as $n \to \infty$, in other words if we let the mesh size of the partition tend to zero. Accordingly, we will call Brownian motion a solution of the equation $U'(t) =$'noise', $U(0) = 0$. The 'noise' that is thus described is called *white noise*. Actually, we have only described the anti-derivative of white noise. To obtain the noise itself we only have to differentiate. However, almost every path of the Brownian motion is nowhere differentiable. Therefore we will consider integral equations rather than differential equations.

The solution of a stochastic integral equation will be a family $(U(t))_{t\geqslant 0}$ of random variables. It is sometimes useful to see every $U(t)$ as an element of a function space. For a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ we will consider the space $L^2(\mathbb{P})$ consisting of (equivalence classes of) measurable real functions f such that

$$\int_\Omega f^2 \, d\mathbb{P} < \infty.$$

Random variables on Ω with finite variance can be interpreted as elements of $L^2(\mathbb{P})$. Indeed, for a random variable X on Ω we have

$$\mathbb{E}X^2 = \int_\Omega X(\omega)^2 \, d\mathbb{P}(\omega) = \|X\|^2_{L^2(\mathbb{P})}.$$

$L^2(\mathbb{P})$ is a Hilbert space with the inner product

$$\langle X, Y \rangle = \int_\Omega X(\omega)Y(\omega) \, d\mathbb{P}(\omega) = \mathbb{E}XY, \quad X, Y \in L^2(\mathbb{P}).$$

All Gaussian random variables on Ω are members of $L^2(\mathbb{P})$.

Exercise 1.33. Show that two Gaussian random variables, one of them centered, on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ are independent if and only if they are orthogonal in the sense of the inner product of $L^2(\mathbb{P})$.

Chapter 2

The Wiener integral

2.1 A model problem

Let us consider the following initial value problem:

$$u'(t) = -au(t) + f(t), \qquad u(0) = u_0,$$

where, for the sake of simplicity, we assume that $f : [0, \infty) \to \mathbb{R}$ is continuous. This problem is equivalent to the integral equation

$$u(t) = u_0 - a \int_0^t u(s)\, ds + F(t),$$

where $F(t) = \int_0^t f(s)\, ds$. By a direct computation involving Fubini's theorem one checks that this integral equation is solved by the function

$$u(t) \;=\; e^{-at} u_0 + \int_0^t e^{-a(t-s)} f(s)\, ds \;=\; e^{-at} u_0 + \int_0^t e^{-a(t-s)}\, dF(s), \tag{2.1}$$

where the right hand side integral is interpreted as a Riemann-Stieltjes integral. By analogy, one could hope to solve the stochastic integral equation

$$U(t) = U_0 - a \int_0^t U(s)\, ds + W(t), \tag{2.2}$$

where $\{W(t)\}_{t \geqslant 0}$ is a Brownian motion, by

$$U(t) = e^{-at} U_0 + \int_0^t e^{-a(t-s)}\, dW(s). \tag{2.3}$$

Of course one has to give a meaning to integral on the right hand side. As it turns out, Brownian motion has bounded *quadratic* variation, but if paths have unbounded variation almost surely. For this reason we will use L^2-techniques to give a rigorous intepretation of the integral in (2.3) rather than attempt to define the integral path-by-path as Riemann-Stieltjes integrals.

2.2 Quadratic variation of Brownian motion

In this section we shall prove that Brownian motion has finite quadratic variation and that its typical paths have unbounded variation.

Lemma 2.1. *Let Y be a centred Gaussian random variable with $\mathbb{E}(Y^2) = s$. Then $\mathbb{E}(Y^4) = 3s^2$.*

Proof: By differentiating under the integral sign we obtain

$$\mathbb{E}(Y^4) = \frac{d^4}{d\lambda^4}\bigg|_{\lambda=0} \mathbb{E}\big(\exp(-i\lambda Y)\big) = \mathbb{E}\left(\frac{d^4}{d\lambda^4}\bigg|_{\lambda=0} \exp(-i\lambda Y)\right) = \frac{d^4}{d\lambda^4}\bigg|_{\lambda=0} \exp(-\tfrac{1}{2}\lambda^2 s) = 3s^2.$$

□

Let $\{W(t)\}_{t\geqslant 0}$ be a Brownian notion on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $T \geqslant 0$ be fixed. For $n = 1, 2, 3, \ldots$ and $j = 0, \ldots, n-1$ let $\Delta_j^n W := W\big(\frac{(j+1)T}{n}\big) - W\big(\frac{jT}{n}\big)$. We think of $\Delta_j^n W$ as the increments of $W(t)$ relative to the partitions $P_n = \left\{0, \frac{T}{n}, \ldots, \frac{(n-1)T}{n}, T\right\}$ of $[0, T]$. The following result says that these increments have finite, nonzero *quadratic* variation relative to these partitions:

Theorem 2.2. *In $L^2(\mathbb{P})$ we have* $\displaystyle\lim_{n\to\infty}\sum_{j=0}^{n-1}(\Delta_j^n W)^2 = T.$

Proof: From

$$\mathbb{E}(\Delta_j^n W) = 0, \quad \mathbb{E}(\Delta_j^n W)^2 = \frac{T}{n}, \quad \mathbb{E}(\Delta_j^n W)^4 = \frac{3T^2}{n^2},$$

and the independence of the increments $\Delta_j^n W$ it follows that

$$\begin{aligned}
\left\|\sum_{j=0}^{n-1}(\Delta_j^n W)^2 - T\right\|_{L^2(\mathbb{P})}^2 &= \mathbb{E}\left(\left[\left(\sum_{j=0}^{n-1}(\Delta_j^n W)^2\right) - T\right]^2\right) \\
&= \mathbb{E}\left(\left[\sum_{j=0}^{n-1}\left((\Delta_j^n W)^2 - \frac{T}{n}\right)\right]^2\right) \\
&= \mathbb{E}\left(\sum_{j=0}^{n-1}\left((\Delta_j^n W)^2 - \frac{T}{n}\right)^2\right. \\
&\qquad\left. + 2\sum_{0\leqslant j<k\leqslant n-1}\left((\Delta_j^n W)^2 - \frac{T}{n}\right)\left((\Delta_k^n W)^2 - \frac{T}{n}\right)\right) \\
&= \sum_{j=0}^{n-1}\mathbb{E}\left[\left((\Delta_j^n W)^2 - \frac{T}{n}\right)^2\right] \\
&\qquad + 2\sum_{0\leqslant j<k\leqslant n-1}\mathbb{E}\left((\Delta_j^n W)^2 - \frac{T}{n}\right)\mathbb{E}\left((\Delta_k^n W)^2 - \frac{T}{n}\right) \\
&= \sum_{j=0}^{n-1}\mathbb{E}\left[\left((\Delta_j^n W)^2 - \frac{T}{n}\right)^2\right]
\end{aligned}$$

$$= \sum_{j=0}^{n-1} \left[\mathbb{E}\left((\Delta_j^n W)^4\right) - \frac{2T}{n}\mathbb{E}\left((\Delta_j^n W)^2\right) + \frac{T^2}{n^2} \right] = \sum_{j=0}^{n-1} \left(\frac{3T^2}{n^2} - \frac{2T^2}{n^2} + \frac{T^2}{n^2} \right) = \frac{2T^2}{n}.$$

Passing to the limit $n \to \infty$ gives the desired result. □

From this theorem we will deduce next that the typical path of a Brownian motion has unbounded variation.

Proposition 2.3. *Suppose $w : [a, b] \to \mathbb{R}$ is continuous. Let $\{P_n\}_{n \geqslant 0}$ be a sequence of partitions of $[a, b]$ such that $\lim_{n\to\infty} m(P_n) = 0$. For each n we write $P_n = \{t_0^n, \ldots, t_{N_n}^n\}$. If*

$$\limsup_{n\to\infty} \sum_{j=0}^{N_n-1} |w(t_{j+1}^n) - w(t_j^n)| < \infty,$$

then

$$\lim_{n\to\infty} \sum_{j=0}^{N_n-1} \left|w(t_{j+1}^n) - w(t_j^n)\right|^2 = 0.$$

Proof: Trivially, for all n we have

$$\sum_{j=0}^{N_n-1} \left|w(t_{j+1}^n) - w(t_j^n)\right|^2 \leqslant \left(\max_{j=0,\ldots,N_n-1} |w(t_{j+1}^n) - w(t_j^n)| \right) \cdot \sum_{j=0}^{N_n-1} |w(t_{j+1}^n) - w(t_j^n)|. \quad (2.4)$$

Since w is continuous we have $\lim_{n\to\infty} \left(\max_{j=0,\ldots,N_n-1} |w(t_{j+1}^n) - w(t_j^n)| \right) = 0$, so by taking the limeas superior on both sides in (2.4) we obtain $\lim_{n\to\infty} \left(\sum_{j=0}^{N_n-1} \left|w(t_{j+1}^n) - w(t_j^n)\right|^2 \right) = 0.$ □

Corollary 2.4. *Almost surely, the paths of a Brownian motion have unbounded variation.*

Proof: Passing to an almost surely convergent subsequence in Theorem 2.2, we obtain

$$\lim_{k\to\infty} \sum_{j=0}^{n_k} \big(W(t_{j+1}^{n_k}) - W(t_j^{n_k})\big)^2 = \lim_{n\to\infty} \sum_{j=0}^{n_k} (\Delta_j^{n_k} W)^2 = T \qquad \text{almost surely.}$$

In particular this limit is nonzero almost surely, and we can apply Proposition 2.3. □

2.3 The Wiener integral

If $f : [0, T] \to \mathbb{R}$ is a simple function of the form $f = \sum_{j=0}^{N-1} c_j \mathbf{1}_{[t_j, t_{j+1})}$, with $0 \leqslant t_0 < \cdots < t_N \leqslant T$, then its Riemann-Stieltjes integral is given by

$$\int_0^T f(t)\, dw(t) = \sum_{j=0}^{N-1} c_j \big(w(t_{j+1}) - w(t_j)\big).$$

Analogously, for simple functions f we can define a *random variable*, which we denote by $f \cdot W$, by the formula

$$f \cdot W := \sum_{j=0}^{N-1} c_j \big(W(t_{j+1}) - W(t_j)\big).$$

Clearly, this analogy suggest the notation $f \cdot W = \int_0^T f(t)\, dW(t)$, but for the moment we shall resist this temptation.
Let us try instead to compute the $L^2(\mathbb{P})$-norm of $f \cdot W$. We have

$$(f \cdot W)^2 = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} c_j c_k (\Delta_j W)(\Delta_k W) = \sum_{j=0}^{N-1} c_j^2 (\Delta_j W)^2 + 2 \sum_{0 \leqslant j < k \leqslant N-1} c_j c_k (\Delta_j W)(\Delta_k W).$$

Now

$$\mathbb{E}\big(c_j^2 (\Delta_j W)^2\big) = c_j^2 (t_{j+1} - t_j).$$

If $0 \leqslant j < k \leqslant N-1$, then $c_j c_k \Delta_j W$ and $\Delta_k W$ are independent, and therefore, recalling that $\mathbb{E}(\Delta_k W) = 0$,

$$\mathbb{E}\big(c_j c_k (\Delta_j W)(\Delta_k W)\big) = \mathbb{E}(c_j c_k \Delta_j W)\mathbb{E}(\Delta_k W) = 0.$$

Putting this together we obtain

$$\|f \cdot W\|_{L^2(\mathbb{P})}^2 = \mathbb{E}\big((f \cdot W)^2\big) = \sum_{j=0}^{N-1} c_j^2 (t_{j+1} - t_j) = \int_0^T |f(t)|^2\, dt = \|f\|_{L^2[0,T]}^2.$$

Since the simple functions are dense in $L^2[0,T]$, we have proved:

Theorem 2.5. *The map $I_W : f \mapsto f \cdot W$ extends uniquely to an isometry I_W from $L^2[0,T]$ into $L^2(\mathbb{P})$.*

We call $I_W(f)$ the *Wiener integral* of $f \in L^2[0,T]$ with respect to $\{W(t)\}_{t \geqslant 0}$; notation:

$$\int_0^T f(t)\, dW(t).$$

- *Exercise.* Show that $\int_0^T t\, dW(t) = TW(T) - \int_0^T W(t)\, dt$.
- *Exercise.* Show that for all $0 \leqslant t \leqslant T$ and $f \in L^2[0,T]$ we have

$$\int_0^t f(s)\, dW(s) = \int_0^T \mathbf{1}_{[0,t]}(s) f(s)\, dW(s).$$

- *Exercise.* Show that for all $f, g \in L^2[0,T]$ we have

$$\mathbb{E}\left[\left(\int_0^T f(t)\, dW(t)\right)\left(\int_0^T g(t)\, dW(t)\right)\right] = \int_0^T f(t) g(t)\, dt. \tag{2.5}$$

Theorem 2.6. *For all $f \in L^2[0,T]$ the random variable $I_W(f)$ is centred Gaussian, with*

$$\mathbb{E}\big((I_W(f))^2\big) = \|f\|_{L^2[0,T]}^2.$$

Proof: The identity (2.5) is a reformulation of Theorem 2.5. It remains to show that $I_W(f)$ is centred Gaussian. Since $I_W(f)$ is centred Gaussian for stepfunctions f, it suffices to show that the $L^2(\mathbb{P})$-limit g of a sequence $\{g_n\}_{n\geqslant 0}$ of centred Gaussian random variables is again centred Gaussian.
Let μ denote the distribution of g, i.e. the probability measure on $\mathbb{R}$ given by

$$\mu(A) = \mathbb{P}\big(\omega \in \Omega : \ g(\omega) \in A\big).$$

Let us denote by $\hat{\mu}$ its Fourier transform:

$$\hat{\mu}(\xi) = \int_{-\infty}^{\infty} \exp(-i\xi t)\, d\mu(t) = \mathbb{E}\big(\exp(-i\xi g)\big), \qquad \xi \in \mathbb{R}.$$

We define μ_n and $\hat{\mu}_n$ similarly. Since μ_n is a centred Gaussian distribution, there exists $s_n \geqslant 0$ such that $\hat{\mu}_n(\xi) = \exp(-\frac{1}{2}s_n\xi^2)$ for all $\xi \in \mathbb{R}$. For all $\xi \in \mathbb{R}$ we have

$$\lim_{n\to\infty} \hat{\mu}_n(\xi) = \lim_{n\to\infty} \mathbb{E}\big(\exp(-i\xi g_n)\big) = \mathbb{E}\big(\exp(-i\xi g)\big) = \hat{\mu}(\xi).$$

The function $\xi \mapsto \hat{\mu}(\xi)$ is continuous and takes the value $\hat{\mu}(0) = 1$. Hence we can fix $\xi > 0$ so small that $\hat{\mu}(\xi) \neq 0$. Then,

$$\lim_{n\to\infty} e^{-\frac{1}{2}s_n\xi^2} = \lim_{n\to\infty} \hat{\mu}_n(\xi) = \hat{\mu}(\xi) \neq 0.$$

From this it follows that $s := \lim_{n\to\infty} s_n$ exists. Then,

$$\hat{\mu}(\xi) = \lim_{n\to\infty} \hat{\mu}_n(\xi) = \lim_{n\to\infty} e^{-\frac{1}{2}s_n\xi^2} = e^{-\frac{1}{2}s\xi^2}, \qquad \xi \in \mathbb{R}.$$

But this shows that μ is a centred Gaussian distribution, so g is centred Gaussian. $\square$

- *Exercise.* In the proof we used the following observation: if $\lim_{n\to\infty} g_n = g$ in $L^2(\mathbb{P})$, then for all $\xi \in \mathbb{R}$ we have $\lim_{n\to\infty} \mathbb{E}\big(\exp(-i\xi g_n)\big) = \mathbb{E}\big(\exp(-i\xi g)\big)$. Prove this.

2.4 The stochastic Fubini theorem

Our next aim is to prove a Fubini theorem for the Wiener integral. Before we can state the result we recall some terminology concerning vector-valued integration.

Let μ be a positive σ-finite measure on a measurable space $(A, \mathscr{A})$ and let X be a Banach space. An X-valued *stepfunction* is a function of the form $f = \sum_{j=0}^{N-1} x_j \mathbf{1}_{A_j}$ with $x_j \in X$ and $A_j \in \mathscr{A}$ disjoint with $\mu(A_j) < \infty$ for all j. A function $f : A \to X$ is said to be *strongly measurable* if it is the μ-a.e. pointwise limit of a sequence of stepfunctions. Note that this concept of measurability depends on the measure μ. It is easy to see that if f is strongly measurable, then also the scalar-valued function $\|f\|$ is measurable. We say that f is *Bochner integrable* if f is strongly measurable and

$$\int_A \|f\|\, d\mu < \infty.$$

Clearly every stepfunction is Bochner integrable. For a stepfunction $f = \sum_{j=0}^{N-1} x_j \mathbf{1}_{A_j}$ we define a vector $I(f) \in X$ by

$$I(f) := \sum_{j=0}^{N-1} \mu(A_j)\, x_j.$$

Note that $I(f)$ does not depend upon the representation of f as a stepfunction. If f is an arbitrary Bochner integrable function, it is not hard to show that there exists a sequence of stepfunctions for which

$$\lim_{n\to\infty} \int_A \|f_n - f\|\, d\mu = 0.$$

- *Exercise.* Prove this. *Hint:* For a stepfunction f_n, consider $g_n := f_n \mathbf{1}_{\{\|f_n\| \leqslant 2\|f\|\}}$.

It follows from this that $\{I(f_n)\}_{n\geqslant 0}$ is a Cauchy sequence in X; its limit is denoted by $I(f)$ and is called the *Bochner integral* of f; notation:

$$I(f) := \int_A f\, d\mu.$$

It is easy to show that $I(f)$ does not depend upon the approximating sequence of stepfunctions $\{f_n\}_{n\geqslant 0}$.

- *Exercise.* Show that $\left\|\int_A f\, d\mu\right\| \leqslant \int_A \|f\|\, d\mu$.

- *Exercise.* Let Y be a closed subspace of X. Suppose $f : A \to X$ is Bochner integrable and takes its values in Y for μ-almost all $a \in A$. Prove that $\int_A f\, d\mu \in Y$.

Let $p \in [1, \infty)$. We call a function $f : A \to X$ *p-Bochner integrable* if it is strongly measurable and

$$\int_A \|f\|^p\, d\mu < \infty.$$

The space $L^p((A,\mu);X)$ of all p-Bochner integrable functions is a Banach space with respect to the norm $\|\cdot\|_{L^p((A,\mu);X)}$ defined by

$$\|f\|^p_{L^p((A,\mu);X)} := \int_A \|f\|^p\, d\mu.$$

Lemma 2.7. *If $f \in L^2([0,T]\times[0,T])$, then for almost all $t \in [0,T]$ we have $f(\cdot,t) \in L^2[0,T]$, and the resulting $L^2(\mathbb{P})$-valued function*

$$f^W(t) := \int_0^T f(s,t)\, dW(s), \qquad \text{a.a. } t \in [0,T]$$

is square Bochner integrable, and we have

$$\int_0^T \|f^W(t)\|^2_{L^2(\mathbb{P})}\, dt = \|f\|^2_{L^2([0,T]\times[0,T])}.$$

In particular, $t \mapsto f^W(t)$ is Bochner integrable and

$$\left\|\int_0^T f^W(t)\, dt\right\|_{L^2(\mathbb{P})} \leqslant T^{\frac{1}{2}} \|f\|_{L^2([0,T]\times[0,T])}.$$

Proof: The fact that $f(\cdot,t) \in L^2[0,T]$ for almost all $t \in [0,T]$ is part of Fubini's theorem. Next we prove that f^W is strongly measurable. Choose a sequence $\{f_n\}_{n\geqslant 0}$ of simple functions of the form

$$f_n(s,t) = \sum_{j=0}^{N_n}\sum_{k=0}^{M_n} c_{jk}^n \mathbf{1}_{[t_j^n,t_{j+1}^n)\times[t_k^n,t_{k+1}^n)}(s,t)$$

such that $\lim_{n\to\infty} f_n = f$ in $L^2([0,T]\times[0,T])$. For all $t \in [0,T]$ the function $\phi_n(t) := f_n(\cdot,t)$ belongs to $L^2[0,T]$. By Fubini's theorem,

$$\lim_{n\to\infty} \int_0^T \|\phi_n(t) - f(\cdot,t)\|_{L^2[0,T]}^2\,dt = 0,$$

and therefore we may assume, after passing to a pointwise a.e. convergent subsequence, that for almost all $t \in [0,T]$ we have

$$f(\cdot,t) = \lim_{n\to\infty} \phi_n(t) \quad \text{in } L^2[0,T].$$

It follows that for almost all $t \in [0,T]$ we have

$$f^W(t) = I_W\big(f(\cdot,t)\big) = \lim_{n\to\infty} I_W\big(\phi_n(t)\big) \quad \text{in } L^2(\mathbb{P}).$$

But,

$$I_W\big(\phi_n(t)\big) = \phi_n(t)\cdot W = \sum_{k=0}^{M_n}\left(\sum_{j=0}^{N_n} c_{jk}^n\big(W(t_{j+1}^n) - W(t_j)\big)\right)\mathbf{1}_{[t_k^n,t_{k+1}^n)}(t),$$

which shows that the $L^2(\mathbb{P})$-valued function $t \mapsto I_W\big(\phi_n(t)\big)$ is a stepfunction.
We have shown that f^W is a.e. the limit of a sequence of $L^2(\mathbb{P})$-valued stepfunctions, and therefore f^W is strongly measurable. The square Bochner integrability of f^W now follows from

$$\begin{aligned}\int_0^T \big\|f^W(t)\big\|_{L^2(\mathbb{P})}^2\,dt &= \int_0^T \left\|\int_0^T f(s,t)\,dW(s)\right\|_{L^2(\mathbb{P})}^2 dt \\ &= \int_0^T \|f(\cdot,t)\|_{L^2[0,T]}^2\,dt = \int_0^T\int_0^T |f(s,t)|^2\,ds\,dt = \|f\|_{L^2([0,T]\times[0,T])}^2.\end{aligned}$$

The final assertion follows from this via the Cauchy-Schwarz inequality. □

Lemma 2.8. *If $f \in L^2([0,T]\times[0,T])$, then the function*

$$F(s) := \int_0^T f(s,t)\,dt, \qquad s \in [0,T],$$

belongs to $L^2[0,T]$ and we have

$$\|F\|_{L^2[0,T]} \leqslant T^{\frac12}\|f\|_{L^2([0,T]\times[0,T])}.$$

Proof. Using the Cauchy-Schwarz inequality we estimate:

$$\int_0^T |F(s)|^2\,ds = \int_0^T \left|\int_0^T f(s,t)\,dt\right|^2 ds \leqslant T\int_0^T\int_0^T |f(s,t)|^2\,dt\,ds.$$

□

Now we are in a position to state and prove the following result:

Theorem 2.9 (Fubini's theorem). *For all* $f \in L^2([0,T] \times [0,T])$ *we have*

$$\int_0^T \left(\int_0^T f(s,t)\, dW(s) \right) dt = \int_0^T \left(\int_0^T f(s,t)\, dt \right) dW(s).$$

Remark 2.10. Thanks to Lemma 2.7, the integral on the left makes sense as a Bochner integral in $L^2(\mathbb{P})$, whereas by Lemma 2.8 the integral on the right makes sense as a Wiener integral.

Proof. The theorem holds for any simple function f of the form

$$f = \sum_{j=0}^{N} \sum_{k=0}^{M} c_{jk} \mathbf{1}_{[t_j,t_{j+1}) \times [t_k,t_{k+1})} \tag{2.6}$$

Indeed, by direct computation both expressions are seen to be equal to

$$\sum_{j=0}^{N} \sum_{k=0}^{M} c_{jk} \left(t_{k+1} - t_k\right)\left(W(t_{j+1}) - W(t_j)\right).$$

Now let $f \in L^2([0,T] \times [0,T])$ be arbitrary and let $\{f_n\}_{n\geqslant 0}$ be a sequence of simple functions with $\lim_{n\to\infty} f_n = f$ in $L^2([0,T] \times [0,T])$. Without loss of generality we may assume that each f_n is of the form (2.6).

Using the notations introduced in Lemmas 2.7 and 2.8 and the estimates proved there, we obtain

$$\begin{aligned}
\int_0^T \left(\int_0^T f(s,t)\, dW(s) \right) dt &= \int_0^T f^W(t)\, dt = \lim_{n\to\infty} \int_0^T f_n^W(t)\, dt \\
&= \lim_{n\to\infty} \int_0^T \left(\int_0^T f_n(s,t)\, dW(s) \right) dt = \lim_{n\to\infty} \int_0^T \left(\int_0^T f_n(s,t)\, dt \right) dW(s) \\
&= \lim_{n\to\infty} \int_0^T F_n(s)\, dW(s) = \int_0^T F(s)\, dW(s) \\
&= \int_0^T \left(\int_0^T f(s,t)\, dt \right) dW(s),
\end{aligned}$$

the convergence being in the sense of $L^2(\mathbb{P})$. □

2.5 L^2-solutions for the stochastic integral equation

Let us now turn to the stochastic integral equation

$$U(t) = u_0 - a \int_0^t U(s)\, ds + W(t), \qquad t \in [0,T], \tag{2.7}$$

where $a \in \mathbb{R}$ is a constant. In most applications, a is a positive constant; this explains why we insist on the minus-sign.

We call a function $U : [0,T] \to L^2(\mathbb{P})$ an L^2-*solution* of (2.7) if U is square Bochner integrable and satisfies (2.7) for all $t \in [0,T]$.

Formally we can differentiate both sides of (2.7) and rewrite the equation in differential form as a *stochastic differential equation*:

$$\begin{aligned} dU(t) &= -aU(t)\,dt + dW(t), \qquad t \in [0,T], \\ U(0) &= u_0. \end{aligned} \tag{2.8}$$

We stress, however, that (2.8) has no independent meaning by itself; its rigorous interpretation is (2.7).

Theorem 2.11. *The function $U : [0,T] \to L^2(\mathbb{P})$ defined by*

$$U(t) = e^{-at}u_0 + \int_0^t e^{-a(t-s)}\,dW(s), \qquad t \in [0,T], \tag{2.9}$$

is an L^2-solution of equation (2.7).

Proof: Noting that

$$\int_0^t e^{-a(t-s)}\,dW(s) = \int_0^T \mathbf{1}_{[0,t]}(s)e^{-a(t-s)}\,dW(s),$$

the square Bochner integrability of $U(\cdot)$ on $[0,T]$ follows from Lemma 2.7. Using the stochastic Fubini theorem we obtain

$$\begin{aligned} \int_0^t U(s)\,ds &= \int_0^t \Big(e^{-as}u_0 + \int_0^s e^{-a(s-r)}\,dW(r)\Big)\,ds \\ &= \int_0^t e^{-as}u_0\,ds + \int_0^t \Big(\int_0^t \mathbf{1}_{[0,s]}(r)e^{-a(s-r)}\,dW(r)\Big)\,ds \\ &= \frac{1}{a}\big(1-e^{-at}\big)u_0 + \int_0^t \Big(\int_0^t \mathbf{1}_{[0,s]}(r)e^{-a(s-r)}\,ds\Big)\,dW(r) \\ &= \frac{1}{a}\big(1-e^{-at}\big)u_0 + \int_0^t \Big(\int_r^t e^{-a(s-r)}\,ds\Big)\,dW(r) \\ &= \frac{1}{a}\big(1-e^{-at}\big)u_0 + \frac{1}{a}\int_0^t \Big(1-e^{-a(t-r)}\Big)\,dW(r) \\ &= \frac{1}{a}\big(1-e^{-at}\big)u_0 - \frac{1}{a}\int_0^t e^{-a(t-r)}\,dW(r) + \frac{1}{a}\big(W(t)-W(0)\big) \\ &= \frac{1}{a}\big(1-e^{-at}\big)u_0 - \frac{1}{a}\big(U(t)-e^{-at}u_0\big) + \frac{1}{a}W(t) \\ &= \frac{1}{a}\big(u_0 - U(t) + W(t)\big). \end{aligned}$$

Therefore,

$$u_0 - a\int_0^t U(s)\,ds + W(t) = u_0 - \big(u_0 - U(t) + W(t)\big) + W(t) = U(t),$$

and the theorem is proved. $\square$

• Exercise. Assume that $a > 0$. Show that there exists a centred Gaussian measure μ_∞ on $\mathbb{R}$ such that $\lim_{n\to\infty} \mu_t = \mu_\infty$ weakly, where μ_t is the distribution of the random variable $U(t)$. What does this mean for the equation?

Next we address the question of uniqueness.

Lemma 2.12. *Suppose X is a Banach space and $f : [0,T] \to X$ is a Bochner integrable function satisfying*

$$f(t) = C\int_0^t f(s)\,ds, \qquad t \in [0,T], \tag{2.10}$$

where C is a fixed constant. Then $f(t) = 0$ for all $t \in [0,T]$.

Proof: The right hand side in (2.10) is continous in t, and therefore f is continuous. Hence the right hand side in this identity is continuously differentiable with derivative $Cf(t)$. It follows that the left hand side is continuously differentiable as well, with derivative $f'(t) = Cf(t)$. Since also $f(0) = 0$ it follows that f solves the initial value problem $y' = Cy$, $y(0) = 0$. Since $y = 0$ is the unique solution of this problem, it follows that $f(t) = 0$ for all $t \in [0,T]$. □

• Exercise. Supply the details of this proof.

Theorem 2.13. *The function U defined by* (2.9) *is the unique L^2-solution of the equation* (2.7).

Proof: Suppose $U : [0,T] \to L^2(\mathbb{P})$ and $V : [0,T] \to L^2(\mathbb{P})$ are two L^2-solutions. Then the difference $U - V$ satisfies

$$U(t) - V(t) = -a\int_0^t \big(U(s) - V(s)\big)\,ds, \qquad t \in [0,T].$$

By the lemma we have $U(t) - V(t) = 0$ for all $t \in [0,T]$. □

• Exercise. Let $a \in \mathbb{R}$ and $\sigma \in \mathbb{R}$. Show that the equation

$$U(t) = u_0 - a\int_0^t U(s)\,ds + \sigma W(t), \qquad t \in [0,T],$$

has a unique L^2-solution, and find an explicit formula for it.

2.6 Path regularity

Having obtained a unique L^2-solution $U : [0,T] \to L^2(\mathbb{P})$, we now ask how this solution behaves pathwise. There is a difficulty here: for every fixed $t \in [0,T]$, the random variable $U(t)$ is an element of $L^2(\mathbb{P})$, and as such defined as an *equivalence class* of functions. For this reason, it makes no sense to talk about the 'paths' $t \mapsto (U(t))(\omega) =: U(t,\omega)$. In order to overcome this difficulty we define $\mathscr{L}^2(\mathbb{P})$ as the vector space of all square integrable functions $f : \Omega \to \mathbb{R}$. Here we do *not* identify functions that are equal $\mathbb{P}$-a.e. We will call the function $\hat{U} : [0,T] \to \mathscr{L}^2(\mathbb{P})$ a *version* of U if for every $t \in [0,T]$ the function $\hat{U}(t)$ belongs to the equivalence class of $U(t)$.

Theorem 2.14. *The solution of equation* (2.7) *admits a version whose paths are continuous $\mathbb{P}$-almost surely.*

Proof: We begin by recalling that the solution $U : [0,T] \to L^2(\mathbb{P})$ is strongly measurable. Therefore we can find a sequence of stepfunctions $U_n : [0,T] \to L^2(\mathbb{P})$ such that $\lim_{n\to\infty} U_n = U$ in $L^2([0,T]; L^2(\mathbb{P}))$. For every n we choose a representing stepfunction $\tilde{U}_n : [0,T] \to \mathscr{L}^2(\mathbb{P})$. Then for every n, the function $\tilde{V}_n : [0,T] \times \Omega$ defined by

$$\tilde{V}_n(t,\omega) := (\tilde{U}_n(t))(\omega), \qquad \omega \in \Omega,$$

is joint measurable. Let $V_n \in L^2(m_T \times \mathbb{P})$ denote the equivalence class of $\tilde{V}_n$; here m_T denotes the Lebesgue measure of $[0,T]$. For all n, m we have

$$\|V_n - V_m\|^2_{L^2(m_T\times\mathbb{P})} = \int_0^T \|U_n(t) - U_m(t)\|^2_{L^2(\mathbb{P})}\, dt = \|U_n - U_m\|^2_{L^2([0,T];L^2(\mathbb{P}))}.$$

By Fubini's theorem it follows that $\{V_n\}_{n\geqslant 0}$ is a Cauchy sequence in $L^2(m_T \times \mathbb{P})$. Let $X \in L^2(m_T \times \mathbb{P})$ denote the limit. By Fubini's theorem, both $X \in L^2([0,T]; L^2(\mathbb{P}))$ and $V_n \in L^2([0,T]; L^2(\mathbb{P}))$ for all n, and in $L^2([0,T]; L^2(\mathbb{P}))$ we have

$$U = \lim_{n\to\infty} U_n = \lim_{n\to\infty} V_n = X.$$

In particular, for almost all $t \in [0,T]$ we have $U(t) = X(t,\cdot)$ in $L^2(\mathbb{P})$.
Let $\hat{X} : [0,T] \times \Omega \to \mathbb{R}$ be a joint measurable function in the equivalence class of X. Then by Fubini's theorem, for almost all $\omega \in \Omega$ the function $t \mapsto \hat{X}(t,\omega)$ is measurable and square integrable. Also, for almost all $\omega \in \Omega$ the path $t \mapsto W(t,\omega)$ is continuous. For those $\omega \in \Omega$ for which both assertions hold we define

$$\hat{U}(t,\omega) := u_0 + \int_0^t \hat{X}(s,\omega)\, ds + W(t,\omega), \qquad t \in [0,T].$$

Clearly, $t \mapsto \hat{U}(t,\omega)$ is continuous. For the remaining $\omega \in \Omega$ we put $\hat{U}(t,\omega) := 0$. By Fubini's theorem, $\hat{U}(t) := \hat{U}(t,\cdot) \in \mathscr{L}^2(\mathbb{P})$ for all $t \in [0,T]$. The proof is complete if we show that for all $t \in [0,T]$, the function $\hat{U}(t)$ belongs to the equivalence class of $U(t)$.
Let $\overline{U}(t)$ denote the equivalence class of $\hat{U}(t)$; we have to show that $\overline{U}(t) = U(t)$. Let $\hat{Y} \in \mathscr{L}^2(\mathbb{P})$ be arbitrary and let $Y \in L^2(\mathbb{P})$ be its equivalence class. Then,

$$\begin{aligned}
\mathbb{E}(\overline{U}(t), Y) = \mathbb{E}(\hat{U}(t)\hat{Y}) &= u_0\mathbb{E}(\hat{Y}) - a\,\mathbb{E}\int_0^t \hat{X}(s)\hat{Y}\, ds + \mathbb{E}(W(t)Y) \\
&= u_0\mathbb{E}(\hat{Y}) - a\int_0^t \mathbb{E}(\hat{X}(s)\hat{Y})\, ds + \mathbb{E}(W(t)Y) \\
&= u_0\mathbb{E}(Y) - a\int_0^t \mathbb{E}(X(s)Y)\, ds + \mathbb{E}(W(t)Y) \\
&= u_0\mathbb{E}(Y) - a\int_0^t \mathbb{E}(U(s)Y)\, ds + \mathbb{E}(W(t)Y) \\
&= u_0\mathbb{E}(Y) - a\,\mathbb{E}\int_0^t U(s)Y\, ds + \mathbb{E}(W(t)Y) \\
&= \mathbb{E}(U(t)Y).
\end{aligned}$$

Hence by the Hahn-Banach theorem, $\overline{U}(t) = U(t)$. □

Chapter 3

The Itô integral

3.1 Construction of the Itô integral

For a simple function $f : [0,T] \to \mathbb{R}$ of the form

$$f = \sum_{j=0}^{N-1} c_j \mathbf{1}_{[t_j,t_{j+1})}$$

we have defined the Wiener integral by

$$f \cdot W := \sum_{j=0}^{N-1} c_j \big(W(t_{j+1}) - W(t_j)\big).$$

The identity $\|f \cdot W\|^2_{L^2(\mathbb{P})} = \|f\|^2_{L^2[0,T]}$ enabled us to extend the integral to arbitary functions $f \in L^2[0,T]$.
Let us now suppose that $g : [0,T] \to L^2(\mathbb{P})$ is a simple function of the form

$$g = \sum_{j=0}^{N-1} \eta_j \mathbf{1}_{[t_j,t_{j+1})}, \tag{3.1}$$

where the $\eta_j \in L^2(\mathbb{P})$ are given random variables. Then it makes sense to define a random variable $g \cdot W$ by

$$g \cdot W := \sum_{j=0}^{N-1} \eta_j \big(W(t_{j+1}) - W(t_j)\big).$$

It is not obvious, however, whether $g \cdot W$ belongs to $L^2(\mathbb{P})$ again and whether this formula can be extended to some reasonable class of $L^2(\mathbb{P})$-valued functions. It turns out that the answers to both questions are affirmative for *adapted* functions g. Before defining the concept of adaptedness, we first need to discuss *filtrations*.

Suppose $(\Omega, \mathscr{F})$ is a measurable space. A *filtration* of $(\Omega, \mathscr{F})$ is a one-parameter family $\{\mathscr{F}_t\}_{t\geqslant 0}$ of sub-σ-fields of $\mathscr{F}$ with the property that

$$\mathscr{F}_s \subseteq \mathscr{F}_t \quad \text{for all } 0 \leqslant s \leqslant t.$$

If one interprets the elements of $\mathscr{F}$ as the set of all possible 'events', then the usual interpertation of $\mathscr{F}_t$ is the set of all possible 'events up to time t'. Thus, $\mathscr{F}_t$ can be thought of as the sub-σ-field of $\mathscr{F}$ containing all 'information' about the 'past up to time t'.
If $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space with filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$, we can define a one-parameter family of linear subspaces $\{L^2_t(\mathbb{P})\}_{t\geqslant 0}$ of $L^2(\mathbb{P})$ by

$$L^2_t(\mathbb{P}) = \big\{\eta \in L^2(\mathbb{P}) : \ \eta \text{ is } \mathscr{F}_t\text{-measurable}\big\}.$$

Clearly each $L^2_t(\mathbb{P})$ is a *closed* linear subspace of $L^2(\mathbb{P})$, and we have

$$L^2_s(\mathbb{P}) \subseteq L^2_t(\mathbb{P}) \quad \text{for all } 0 \leqslant s \leqslant t.$$

We call a function $g : I \to L^2(\mathbb{P})$, with $I \subseteq [0,\infty)$, *adapted* to the filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$ if $f(t) \in L^2_t(\mathbb{P})$ for all $t \in I$, i.e. if $f(t)$ is $\mathscr{F}_t$-measurable for all $t \in I$. In view of the above interpretation of $\mathscr{F}_t$ as the sub-σ-field of all events up to time t, adapted functions are sometimes called *non-anticipating*. Note that the simple function g defined by (3.1) is adapted if and only if η_j is $\mathscr{F}_{t_j}$-measurable for every $j = 0, \dots, n-1$.

A Brownian motion $\{W(t)\}_{t\geqslant 0}$ on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ will be called *adapted* to a filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$ of $(\Omega, \mathscr{F})$ if each $W(t)$ is $\mathscr{F}_t$-measurable and if the increments $W(t) - W(s)$ are independent of $\mathscr{F}_s$ for all $0 \leqslant s \leqslant t$.
Every Brownian motion $\{W(t)\}_{t\geqslant 0}$ is adapted to its *natural filtration* defined by $\mathscr{F}_t = \sigma\{W(s) : 0 \leqslant s \leqslant t\}$, where the right hand side denotes the σ-filed generated by the random variables $\{W(s) : \ 0 \leqslant s \leqslant t\}$. Sometimes it is desirable to consider richer filtrations, however. Therefore from this point onwards we will always assume that $\{W(t)\}_{t\geqslant 0}$ is a Brownian motion adapted to some given filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$.

Lemma 3.1. *If a simple function $g : [0,T] \to L^2(\mathbb{P})$ is adapted to $\{\mathscr{F}_t\}_{t\geqslant 0}$, then $g \cdot W \in L^2_T(\mathbb{P})$ and*

$$\|g \cdot W\|_{L^2(\mathbb{P})} = \|g\|_{L^2([0,T];L^2(\mathbb{P}))}.$$

Proof: Let $g = \sum_{j=0}^{N-1} \eta_j \chi_{[t_j,t_{j+1})}$ with each η_j being $\mathscr{F}_{t_j}$-measurable. The fact that $g \cdot W$ is $\mathscr{F}_T$-mearurable follows from the adaptedness of g and $\{W(t)\}_{t\geqslant 0}$. Moreover,

$$\begin{aligned}(g \cdot W)^2 &= \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \eta_j \eta_k (\Delta_j W)(\Delta_k W) \\ &= \sum_{j=0}^{N-1} \eta_j^2 (\Delta_j W)^2 + 2 \sum_{0\leqslant j<k\leqslant N-1} \eta_j \eta_k (\Delta_j W)(\Delta_k W).\end{aligned}$$

Now η_j is $\mathscr{F}_{t_j}$-measurable and $\Delta_j W = W(t_{j+1}) - W(t_j)$ is independent of $\mathscr{F}_{t_j}$. Therefore,

$$\mathbb{E}\big(\eta_j^2 (\Delta_j W)^2\big) = \mathbb{E}\big(\eta_j^2)\mathbb{E}(\Delta_j W)^2\big) = \mathbb{E}(\eta_j^2)(t_{j+1} - t_j).$$

For $0 \leqslant j < k \leqslant N-1$, the random variable η_j is $\mathscr{F}_{t_j}$-measurable, hence also $\mathscr{F}_{t_k}$-measurable. Also η_k is $\mathscr{F}_{t_k}$-measurable. The increment $\Delta_j W$ is $\mathscr{F}_{t_{j+1}}$-measurable, hence $\mathscr{F}_{t_k}$ measurable since $j+1 \leqslant k$. It follows that $\eta_j \eta_k \Delta_j W$ is $\mathscr{F}_{t_k}$-measurable. On the other hand the increment $\Delta_k W$ is independent of $\mathscr{F}_{t_k}$. Therefore,

$$\mathbb{E}\big(\eta_j \eta_k (\Delta_j W)(\Delta_k W)\big) = \mathbb{E}(\eta_j \eta_k \Delta_j W)\mathbb{E}(\Delta_k W) = 0,$$

recalling that $\mathbb{E}(\Delta_k W) = 0$. Note that $\eta_j \eta_k \Delta_j W$ is integrable, being the product of the $L^2(\mathbb{P})$-functions $\eta_j \Delta_j W$ and η_k.
Putting this together we obtain

$$\|g \cdot W\|^2_{L^2(\mathbb{P})} = \mathbb{E}\left((g \cdot W)^2\right) = \sum_{j=0}^{N-1} \mathbb{E}(\eta_j^2)(t_{j-1} - t_j) = \int_0^T \mathbb{E}\big(|g(t)|^2\big)\, dt = \|g\|^2_{L^2([0,T];L^2(\mathbb{P}))}.$$

□

• *Exercise.* This exercise illustrates the rôle played by the adapteness assumption. For $n = 1, 2, \ldots$ and $j = 0, \ldots, n$ let $t_j^n = \frac{jT}{n}$. Find the following limits in $L^2(\mathbb{P})$:

$$\lim_{n\to\infty} \sum_{j=0}^{n-1} W(t_j^n)\big(W(t_{j+1}^n) - W(t_j^n)\big),$$
$$\lim_{n\to\infty} \sum_{j=0}^{n-1} W(t_{j+1}^n)\big(W(t_{j+1}^n) - W(t_j^n)\big).$$

Note that, formally, both sums can be interpreted as Riemann sums of the integral $\int_0^T W(t)\, dW(t)$ to be defined below. In the first limit we take the left endpoint of each subinterval as sample point and in the second limit we take the right endpoint.

Hint: Use the identities

$$a(b-a) = \tfrac{1}{2}(b^2 - a^2) - \tfrac{1}{2}(a-b)^2,$$
$$b(b-a) = \tfrac{1}{2}(b^2 - a^2) + \tfrac{1}{2}(a-b)^2.$$

We now define

$$\mathscr{M}_T^2 := \Big\{ f \in L^2([0,T]; L^2(\mathbb{P})) :\ f \text{ is adapted to } \{\mathscr{F}_t\}_{t\geqslant 0} \Big\}.$$

This is a closed linear subspace of the Hilbert space $L^2([0,T]; L^2(\mathbb{P}))$; endowed with the norm of $L^2([0,T]; L^2(\mathbb{P}))$ it is a Hilbert space in its own right. In view of the previous lemma it is a natural question whether the adapted simple functions are dense in $\mathscr{M}_T^2$. The following proposition gives the affirmative answer. In the proof we need the following lemma:

Lemma 3.2. *Let X be a Banach space and let $f \in L^2([0,T]; X)$. Define, for $t \in (0,T)$,*

$$f_t(s) = \begin{cases} g\big(s-t\big), & s \in [t,T], \\ 0, & s \in [0,t). \end{cases}$$

Then,

$$\lim_{t\downarrow} \|g_t - g\|_{L^2([0,T];X)} = 0.$$

Proof. By dominated convergence, this is true for stepfunctions, and the general case follows from this by approximation. □

Proposition 3.3. *For every $g \in \mathscr{M}_T^2$ there exists a sequence $\{g_n\}_{n\geqslant 0}$ of adapted simple functions in $\mathscr{M}_T^2$ with*

$$\lim_{n\to\infty} \|g_n - g\|_{L^2([0,T];L^2(\mathbb{P}))} = 0.$$

Proof: Let $\varepsilon > 0$ be given. By the lemma we can choose $N \geqslant 2$ so large that

$$\|g - g_N\|_{L^2([0,T];L^2(\mathbb{P}))} < \tfrac{1}{2}\varepsilon,$$

where

$$g_N(t) = \begin{cases} g\big(t - \frac{T}{N}\big), & t \in [\frac{T}{N}, T], \\ 0, & t \in [0, \frac{T}{N}). \end{cases}$$

Note that g_N is again adapted.
For each $k = 1, \dots, N-1$ we choose a simple function $\phi_k : [\frac{kT}{N}, \frac{(k+1)T}{N}) \to L^2(\mathbb{P})$ such that

$$\Big\| g_N\big|_{[\frac{kT}{N}, \frac{(k+1)T}{N})} - \phi_k \Big\|_{L^2\left([\frac{kT}{N}, \frac{(k+1)T}{N});L^2(\mathbb{P})\right)} < \frac{\varepsilon}{2(N-1)}.$$

Define $\phi : [0,T] \to L^2(\mathbb{P})$ by gluing together the functions ϕ_k:

$$\phi(t) = \begin{cases} \phi_k(t), & t \in [\frac{kT}{N}, \frac{(k+1)T}{N}),\ k = 1, \dots, N-1, \\ 0, & t \in [0, \frac{T}{N}). \end{cases}$$

For each $k = 1, \dots, N-1$ let P_k denote the orthonormal projection in $L^2(\mathbb{P})$ onto $L^2_{\frac{kT}{N}}(\mathbb{P})$ and define

$$\psi_k(t) := P_k(\phi_k(t)), \qquad t \in [\tfrac{kT}{N}, \tfrac{(k+1)T}{N}).$$

Then the simple function $\psi : [0,T] \to L^2(\mathbb{P})$ obtained by gluing the ψ_k together,

$$\psi(t) = \begin{cases} \psi_k(t), & t \in [\frac{kT}{N}, \frac{(k+1)T}{N}),\ k = 1, \dots, N-1, \\ 0, & t \in [0, \frac{T}{N}), \end{cases}$$

is adapted to the filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$. For each $t \in [\frac{kT}{N}, \frac{(k+1)T}{N})$, the random variable $g_N(t)$ is $\mathscr{F}_{t-\frac{T}{N}}$-measurable, hence certainly $\mathscr{F}_{\frac{kT}{N}}$-measurable. It follows that $g_N(t) = P_k(g_N(t))$ and therefore,

$$\|g_N(t) - \psi_k(t)\|_{L^2(\mathbb{P})} = \|P_k(g_N(t) - \phi_k(t))\|_{L^2(\mathbb{P})} \leqslant \|g_N(t) - \phi_k(t)\|_{L^2(\mathbb{P})}.$$

Integrating over $[0,T]$ gives

$$\|g_N - \psi\|_{L^2([0,T];L^2(\mathbb{P}))} \leqslant \|g_N - \phi\|_{L^2([0,T];L^2(\mathbb{P}))} < (N-1) \cdot \frac{\varepsilon}{2(N-1)} = \tfrac{1}{2}\varepsilon.$$

It follows that

$$\|g - \psi\|_{L^2([0,T];L^2(\mathbb{P}))} \leqslant \|g - g_N\|_{L^2([0,T];L^2(\mathbb{P}))} + \|g_N - \psi\|_{L^2([0,T];L^2(\mathbb{P}))} < \tfrac{1}{2} + \tfrac{1}{2}\varepsilon = \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, the lemma is proved. $\square$

Combination of the two lemmas gives the following result.

Theorem 3.4. *The map $I_W : g \mapsto g \cdot W$ has a unique extension to an isometry I_W from $\mathscr{M}_T^2$ into $L_T^2(\mathbb{P})$.*

- ***Exercise.*** Verify that $W(\cdot) \in \mathscr{M}_T^2$ and that

$$\int_0^T W(t)\, dW(t) = \tfrac{1}{2}((W(T))^2 - \tfrac{1}{2}T.$$

- ***Exercise.*** Verify that $(W(\cdot))^2 \in \mathscr{M}_T^2$ and that

$$\int_0^T (W(t))^2\, dW(t) = \tfrac{1}{3}((W(T))^3 - \int_0^T W(t)\, dt.$$

Use the identity $a^2(b-a) = \frac{1}{3}(b^3 - a^3) - a(b-a)^2 - \frac{1}{3}(b-a)^3$.

- ***Exercise.*** Show that for all $0 \leqslant t \leqslant T$ and $f \in \mathscr{M}_T^2$ we have

$$\int_0^t f(s)\, dW(s) = \int_0^T \mathbf{1}_{[0,t]}(s) f(s)\, dW(s).$$

- ***Exercise.*** Show that for all $f, g \in \mathscr{M}_T^2$ we have

$$\mathbb{E}\left[\left(\int_0^T f(t)\, dW(t)\right)\left(\int_0^T g(t)\, dW(t)\right)\right] = \int_0^T \mathbb{E}\big(f(t)g(t)\big)\, dt.$$

In the next section the following two simple observations will play an important rôle:

Lemma 3.5. *Let $g \in \mathscr{M}_T^2$. Then the function $g^W : [0,T] \to L^2(\mathbb{P})$ defined by*

$$g^W(t) = \int_0^t g(s)\, dW(s), \qquad t \in [0,T],$$

belongs to $\mathscr{M}_T^2$ again.

Proof: It is immediate from the theorem that for every $t \in [0,T]$, the random variable $g^W(t)$ is $\mathscr{F}_t$-measurable. Hence to prove that g^W belongs to $\mathscr{M}_T^2$ it remains to check that g^W is strongly measurable and square integrable.
For $0 \leqslant t < t' \leqslant T$ we have

$$\big\|g^W(t') - g^W(t)\big\|_{L^2(\mathbb{P})} = \left\|\int_0^T \mathbf{1}_{[t,t')}(s) g(s)\, dW(s)\right\|_{L^2(\mathbb{P})} = \|\mathbf{1}_{[t,t')} g\|_{L^2([0,T];L^2(\mathbb{P}))}.$$

By dominated convergence, this shows that g^W is a continuous $L^2(\mathbb{P})$-valued function, and therefore it is certainly strongly measurable. It remains to check square integrability of g^W:

$$\begin{aligned}\int_0^T \|g^W(t)\|_{L^2(\mathbb{P})}^2\, dt &= \int_0^T \left\|\int_0^t g(s)\, dW(s)\right\|_{L^2(\mathbb{P})} dt \\ &= \int_0^T \|\mathbf{1}_{[0,T]} g\|_{L^2([0,T];L^2(\mathbb{P}))}\, dt \leqslant T\|g\|_{L^2([0,T];L^2(\mathbb{P}))}.\end{aligned}$$

□

The proof of the following lemma is left as an exercise.

Lemma 3.6. *For every $g \in \mathscr{M}_T^2$ the function*

$$G(t) := \int_0^t g(s)\,ds, \qquad t \in [0,T],$$

belongs to $\mathscr{M}_T^2$.

3.2 Nonlinear stochastic integral equations

We have seen how we can solve the stochastic integral equation

$$U(t) = u_0 - a\int_0^t U(s)\,ds + W(t), \qquad t \in [0,T], \tag{3.2}$$

explicitly by means of the Wiener integral

$$U(t) = e^{-at}u_0 + \int_0^t e^{-a(t-s)}\,dW(s), \qquad t \in [0,T].$$

In this section we will use the Itô integral to prove existence and uniqueness of solutions for the following more general integral equation:

$$U(t) = \eta_0 + \int_0^t b(U(s))\,ds + \int_0^t \sigma(U(s))\,dW(s), \qquad t \in [0,T], \tag{3.3}$$

where the random variable η_0 and the functions $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ satisfy appropriate assumptions. Equation (3.2) is a special case of (3.3), with $\eta_0(\omega) = u_0$, $b(s) = -as$ and $\sigma(s) = 1$. In differential form, equation (3.3) is usually written as

$$\begin{aligned} dU(t) &= b(U(t))\,dt + \sigma(U(t))\,dW(t), \qquad t \in [0,T], \\ U(0) &= \eta_0. \end{aligned}$$

As in the previous section, we assume that $\{W(t)\}_{t\geqslant 0}$ is a Brownian motion adapted to some given filtration $\{\mathscr{F}_t\}_{t\geqslant 0}$. The precise assumptions we impose on η_0, b and σ are as follows:

Assumption 3.7. The random variable η_0 belongs to $L^2(\mathbb{P})$ and is $\mathscr{F}_0$-measurable. The functions $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are Lipschitz continuous, i.e. there exist constants $L_b \geqslant 0$ and $L_\sigma \geqslant 0$ such that

$$\begin{aligned} |b(x) - b(y)| &\leqslant L_b|x-y| \\ |\sigma(x) - \sigma(y)| &\leqslant L_\sigma|x-y| \end{aligned}$$

for all $x, y \in \mathbb{R}$.

Note that

$$|b(x)| \leqslant |b(0)| + |b(x) - b(0)| \leqslant |b(0)| + L_b|x|$$

and similarly,

$$|\sigma(x)| \leqslant |\sigma(0)| + L_\sigma|x|.$$

Hence if $\eta \in L^2(\mathbb{P})$, then also $b \circ \eta \in L^2(\mathbb{P})$ and $\sigma \circ \eta \in L^2(\mathbb{P})$, and we have

$$\begin{aligned} \|b \circ \eta\|_{L^2(\mathbb{P})} &\leqslant |b(0)| + L_b\|\eta\|_{L^2(\mathbb{P})}, \\ \|\sigma \circ \eta\|_{L^2(\mathbb{P})} &\leqslant |\sigma(0)| + L_\sigma\|\eta\|_{L^2(\mathbb{P})}. \end{aligned}$$

If moreover $\eta \in L_t^2(\mathbb{P})$, then also $b \circ \eta$ and $\sigma \circ \eta$ belong to $L_t^2(\mathbb{P})$. Hence, if $g \in \mathscr{M}_T^2$, then $b \circ g$ and $\sigma \circ g$ belong to $\mathscr{M}_T^2$; in fact,

$$\|b \circ g\|_{L^2([0,T];L^2(\mathbb{P}))} \leqslant T^{\frac{1}{2}}|b(0)| + L_b\|g\|_{L^2([0,T];L^2(\mathbb{P}))},$$
$$\|\sigma \circ g\|_{L^2([0,T];L^2(\mathbb{P}))} \leqslant T^{\frac{1}{2}}|\sigma(0)| + L_\sigma\|g\|_{L^2([0,T];L^2(\mathbb{P}))}.$$

Our proof of existence and uniqueness of solutions will be based on the following classical result:

Theorem 3.8 (Banach Fixed Point Theorem). *Let (X, d) be a complete metric space and let $\Phi : X \to X$ be a strict contraction, i.e. there exists a constant $C \in [0,1)$ such that*

$$d\big(\Phi(x), \Phi(x')\big) \leqslant C\, d(x, x'), \qquad x, x' \in X.$$

Then there exists a unique $x \in X$ with $\Phi(x) = x$.

Proof: Let $x_0 \in X$ be arbitrary and define inductively $x_n := \Phi(x_{n-1})$ for $n \geqslant 1$. From

$$\begin{aligned} d(x_n, x_{n+j}) &\leqslant d(x_n, x_{n+1}) + \cdots + d(x_{n+j-1}, x_{n+j}) \\ &\leqslant C^n d(x_0, x_1) + \cdots + C^{n+j} d(x_0, x_1) \\ &\leqslant \left(\sum_{k=n}^{\infty} C^k\right) d(x_0, x_1) = \frac{C^n}{1-C}\, d(x_0, x_1) \end{aligned}$$

it follows that $\{x_n\}_{n\geqslant 0}$ is a Cauchy sequence. Since (X, d) is complete, this sequence is convergent to a point $x \in X$. Then, noting that Φ is continuous,

$$\Phi(x) = \lim_{n\to\infty} \Phi(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

If also $\Phi(x') = x'$, then

$$d(x, x') = d\big(\Phi(x), \Phi(x')\big) \leqslant C d(x, x'),$$

which is only possible if $d(x, x') = 0$, i.e. if $x = x'$. $\square$

Theorem 3.9. *Under the above assumptions on η_0, b and σ there exists a unique $U \in \mathscr{M}_T^2$ satisfying*

$$U(t) = \eta_0 + \int_0^t b(U(s))\, ds + \int_0^t \sigma(U(s))\, dW(s), \qquad \text{a.a. } t \in [0,T].$$

Proof: We start by noting that if $U \in \mathscr{M}_T^2$, then also $\sigma \circ U \in \mathscr{M}_T^2$ and therefore, by Lemma 3.5, the $L^2(\mathbb{P})$-valued function $t \mapsto \int_0^T \sigma(U(s))\, dW(s)$ belongs to $\mathscr{M}_T^2$. Similarly, from $b \circ U \in \mathscr{M}_T^2$ it follows from Lemma 3.6 that $t \mapsto \int_0^t b(U(s))\, ds$ belongs to $\mathscr{M}_T^2$.
We now define a mapping $\Phi : \mathscr{M}_T^2 \to \mathscr{M}_T^2$ by

$$(\Phi(U))(t) = \eta_0 + \int_0^t b(U(s))\, ds + \int_0^t \sigma(U(s))\, dW(s), \qquad t \in [0,T].$$

Note that $\Phi(U)$ indeed belongs to $\mathscr{M}_T^2$, by the considerations above and the fact that η_0 is $\mathscr{F}_0$-measurable. Our aim is to prove that Φ has a unique fixed point; this fixed point will be the unique

solution within the space $\mathscr{M}_T^2$ of the equation (3.3). With respect to the norm of $\mathscr{M}_T^2$, it is not to be expected that Φ is a contraction, however. Therefore the thing to do is to define an equivalent norm on $\mathscr{M}_T^2$ relative to which Φ turns into a strict contraction.
Define, for $\lambda > 0$ and $U \in \mathscr{M}_T^2$,

$$\|U\|_\lambda^2 := \int_0^T e^{-\lambda t} \|U(t)\|_{L^2(\mathbb{P})}^2 \, dt.$$

Clearly,

$$e^{-\lambda T} \|U\|_{L^2([0,T];L^2(\mathbb{P}))}^2 \leqslant \|U\|_\lambda^2 \leqslant \|U\|_{L^2([0,T];L^2(\mathbb{P}))}^2,$$

so $\|\cdot\|_\lambda$ is an equivalent norm on $\mathscr{M}_T^2$. In particular, the normed space $(\mathscr{M}_T^2, \|\cdot\|_\lambda)$ is complete. For any two $U, V \in \mathscr{M}_T^2$ we have

$$\begin{aligned}
\|\Phi(U) - \Phi(V)\|_\lambda &= \left\| \int_0^\cdot b(U(s)) - b(V(s))\, ds + \int_0^\cdot \sigma(U(s)) - \sigma(V(s))\, dW(s) \right\|_\lambda \\
&\leqslant \left\| \int_0^\cdot b(U(s)) - b(V(s))\, ds \right\|_\lambda + \left\| \int_0^\cdot \sigma(U(s)) - \sigma(V(s))\, dW(s) \right\|_\lambda
\end{aligned}$$

Now,

$$\begin{aligned}
\left\| \int_0^\cdot b(U(s)) - b(V(s))\, ds \right\|_\lambda^2 &= \int_0^T e^{-\lambda t} \left\| \int_0^t b(U(s)) - b(V(s))\, ds \right\|_{L^2(\mathbb{P})}^2 dt \\
&\leqslant \int_0^T e^{-\lambda t} \left(\int_0^t \|b(U(s)) - b(V(s))\|_{L^2(\mathbb{P})}\, ds \right)^2 dt \\
&\leqslant L_b^2 \int_0^T e^{-\lambda t} \left(\int_0^t \|U(s) - V(s)\|_{L^2(\mathbb{P})}\, ds \right)^2 dt \\
&\leqslant L_b^2 \int_0^T e^{-\lambda t} t \int_0^t \|U(s) - V(s)\|_{L^2(\mathbb{P})}^2\, ds\, dt \\
&= L_b^2 \int_0^T \int_0^t \big(e^{-\lambda(t-s)} t\big) \cdot e^{-\lambda s} \|U(s) - V(s)\|_{L^2(\mathbb{P})}^2\, ds\, dt \\
&= L_b^2 \int_0^T \left(\int_s^T e^{-\lambda(t-s)} t\, dt \right) e^{-\lambda s} \|U(s) - V(s)\|_{L^2(\mathbb{P})}^2\, ds \\
&\leqslant L_b^2\, T \int_0^T \left(\int_0^T e^{-\lambda \tau}\, d\tau \right) e^{-\lambda s} \|U(s) - V(s)\|_{L^2(\mathbb{P})}^2\, ds \\
&= \frac{L_b^2 T}{\lambda} \big(1 - e^{-\lambda T}\big) \|U - V\|_\lambda^2 \\
&\leqslant \frac{L_b^2 T}{\lambda} \|U - V\|_\lambda^2.
\end{aligned}$$

If we choose $\lambda_0 \geqslant 16 L_b^2 T$, it follows that

$$\left\| \int_0^\cdot b(U(s)) - b(V(s))\, ds \right\|_\lambda \leqslant \tfrac{1}{4} \|U - V\|_\lambda \qquad \text{for all } \lambda \geqslant \lambda_0.$$

Similarly,

$$
\begin{aligned}
\Big\| \int_0^{\cdot} \sigma(U(s)) - \sigma(V(s))\, dW(s) \Big\|_{\lambda}^2 &= \int_0^T e^{-\lambda t} \Big\| \int_0^t \sigma(U(s)) - \sigma(V(s))\, dW(s) \Big\|_{L^2(\mathbb{P})}^2 dt \\
&= \int_0^T e^{-\lambda t} \int_0^t \|\sigma(U(s)) - \sigma(V(s))\|_{L^2(\mathbb{P})}^2 \, ds\, dt \\
&\leqslant L_\sigma^2 \int_0^T e^{-\lambda t} \int_0^t \|U(s) - V(s)\|_{L^2(\mathbb{P})}^2 \, ds\, dt \\
&= L_\sigma^2 \int_0^T \Big(\int_s^T e^{-\lambda(t-s)}\, ds \Big) e^{-\lambda s} \|U(s) - V(s)\|_{L^2(\mathbb{P})}^2 \, ds \\
&\leqslant \frac{L_\sigma^2}{\lambda} \|U - V\|_\lambda^2.
\end{aligned}
$$

If we choose $\lambda_1 \geqslant 16 L_\sigma^2$, it follows that

$$
\Big\| \int_0^{\cdot} \sigma(U(s)) - \sigma(V(s))\, dW(s) \Big\|_{\lambda} \leqslant \tfrac{1}{4} \|U - V\|_\lambda \qquad \text{for all } \lambda \geqslant \lambda_1.
$$

Hence for all $\lambda \geqslant \max\{\lambda_0, \lambda_1\}$ we obtain

$$
\|\Phi(U) - \Phi(V)\|_\lambda \leqslant \tfrac{1}{4}\|U - V\|_\lambda + \tfrac{1}{4}\|U - V\|_\lambda = \tfrac{1}{2}\|U - V\|_\lambda.
$$

For any such λ we see that Φ is a strict contraction in $(\mathscr{M}_T^2, \|\cdot\|_\lambda)$. Therefore Φ has a unique fixed point by the Banach fixed point theorem. □

Let us call a function $U : [0,T] \to L^2(\mathbb{P})$ an *adapted L^2-solution* of (3.3) if U is adapted, strongly measurable and square Bochner integrable on $[0,T]$ and (3.3) holds for all $t \in [0,T]$.

Theorem 3.10. *Under the above assumptions on η_0, b and σ, equation* (3.3) *has a unique adapted L^2-solution. This solution has a version whose paths are continuous almost surely.*

Proof: Let us define

$$
\tilde{U}(t) := \eta_0 + \int_0^t b(U(s))\, ds + \int_0^t \sigma(U(s))\, dW(s), \qquad t \in [0,T],
$$

where $U \in \mathscr{M}_T^2$ is the function from the theorem. Clearly $\tilde{U}$ is adapted, and for almost all $t \in [0,T]$ the function $\tilde{U}(t)$ belongs to the equivalence class of $U(t)$. In particular, $\tilde{U}$ is strongly measurable and square Bochner integrable on $[0,T]$. It also follows that for all $t \in [0,T]$ we have

$$
\tilde{U}(t) := \eta_0 + \int_0^t b(\tilde{U}(s))\, ds + \int_0^t \sigma(\tilde{U}(s))\, dW(s).
$$

This shows that $\tilde{U}$ is an adapted L^2-solution of (3.3).
Suppose that $\tilde{V}$ is another adapted L^2-solution. Then the uniqueness part of Theorem 3.9 shows that $\tilde{U}(t) = \tilde{V}(t)$ for almost all $t \in [0,T]$. In particular, for all $t \in [0,T]$ we have

$$
\int_0^t b(\tilde{U}(s))\, ds = \int_0^t b(\tilde{V}(s))\, ds \qquad \text{and} \qquad \int_0^t \sigma(\tilde{U}(s))\, dW(s) = \int_0^t \sigma(\tilde{V}(s))\, dW(s).
$$

It follows that $\tilde{U}(t) = \tilde{V}(t)$ holds for all $t \in [0,T]$. This proves uniqueness.
Finally, by repeating the arguments leading to Theorem 2.14 we obtain a version of $\tilde{U}$ with continuous paths almost surely; we leave the details as an exercise. □

3.3 Itô's formula

The existence and uniqueness result proved in the previous section has little practical value, as it gives no hint how to find solutions explicitly. Consider for example the equation

$$U(t) = \eta_0 + \int_0^t U(t)\, dW(t). \tag{3.4}$$

In differential notation it reads

$$dU(t) = U(t)\, dW(t), \qquad U(0) = \eta_0. \tag{3.5}$$

As yet, we have no tools to solve it. Equation (3.5) is the stochastic analoge of the differential equation

$$u'(t) = u(t), \qquad u(0) = u_0.$$

In the case of ordinary differential equations it is sometimes possible to find a solution by integrating the equation in one way or the other. If we want to solve stochastic differential equations in this way, we need an analogue of this procedure.

We first recall the Fundamental Theorem of Calculus. If $F : [0,T] \to \mathbb{R}$ is continuously differentiable, then

$$F(t) - F(0) = \int_0^t F'(s)\, ds, \qquad t \in [0,T].$$

More generally, if $F : [0,T] \times \mathbb{R} \to \mathbb{R}$ has continuous partial derivatives $\frac{\partial F}{\partial t}$ and $\frac{\partial F}{\partial x}$ on $[0,T] \times \mathbb{R}$ and $w : [0,T] \to \mathbb{R}$ is continuously differentiable, then

$$\begin{aligned} F(t, w(t)) - F(0, w(0)) &= \int_0^t \frac{d}{ds} F'\big(s, w(s)\big)\, ds \\ &= \int_0^t \frac{\partial F}{\partial t}\big(s, w(s)\big) + \frac{\partial F}{\partial x}\big(s, w(s)\big) w'(s)\, dt \\ &= \int_0^t \frac{\partial F}{\partial t}\big(s, w(s)\big)\, ds + \int_0^t \frac{\partial F}{\partial x}\big(s, w(s)\big)\, dw(s), \qquad t \in [0,T]. \end{aligned}$$

Itô's formula is the stochastic analogue of this:

Theorem 3.11 (Itô's formula). *Suppose $F : [0,T] \times \mathbb{R} \to \mathbb{R}$ has continuous partial derivatives $\frac{\partial F}{\partial t}$, $\frac{\partial F}{\partial x}$ and $\frac{\partial^2 F}{\partial x^2}$ on $[0,T] \times \mathbb{R}$. Suppose further that $\frac{\partial F}{\partial x}\big(\cdot, W(\cdot)\big) \in \mathscr{M}_T^2$. Then for all $t \in [0,T]$ we have, $\mathbb{P}$-almost surely,*

$$\begin{aligned} &F\big(t, W(t)\big) - F\big(0, W(0)\big) \\ &\quad = \int_0^t \frac{\partial F}{\partial t}\big(s, W(s)\big)\, ds \;+\; \int_0^t \frac{\partial F}{\partial x}\big(s, W(s)\big)\, dW(s) \;+\; \tfrac{1}{2} \int_0^t \frac{\partial^2 F}{\partial x^2}\big(s, W(s)\big)\, ds. \end{aligned}$$

Note the extra term involving the second partial derivative; it will appear in the proof that this term results from the quadradic variation of $\{W(t)\}_{t \geqslant 0}$. Before we prove the theorem, we show how it can be used to compute stochastic integrals and to solve stochastic equations.

Example 3.12. Take $F(t,x) = x^2$. Then

$$\frac{\partial F}{\partial t}(t,x) = 0, \qquad \frac{\partial F}{\partial x}(t,x) = 2x, \qquad \frac{\partial^2 F}{\partial x^2}(t,x) = 2.$$

Therefore by Itô's formula,

$$(W(t))^2 = F(t, W(t)) = \int_0^t 2W(s)\,dW(s) + \tfrac{1}{2}\int_0^t 2\,ds = 2\int_0^t W(s)\,dW(s) + t.$$

It follows that

$$\int_0^t W(s)\,dW(s) = \tfrac{1}{2}(W(t))^2 - \tfrac{1}{2}t.$$

• _Exercise._ Verify that $(W(\cdot))^n \in \mathscr{M}_T^2$ for all $n \geqslant 1$. Use Itô's formula to compute $\int_0^t (W(t))^2\,dW(t)$ and $\int_0^t (W(t))^3\,dW(t)$.

Example 3.13 (Exponential Brownian motion). Let us try to solve (3.4) via Itô's formula. For simplicity we first take $\eta_0 = 1$. It seems reasonable to guess that $U(t) = e^{W(t)}$ is a good candidate for a solution. Let us see what we get if we try to apply Itô's formula, with $F(t,x) = e^x$. Then

$$\frac{\partial F}{\partial t}(t,x) = 0, \qquad \frac{\partial F}{\partial x}(t,x) = \frac{\partial^2 F}{\partial x^2}(t,x) = e^x.$$

In order to apply Itô's formula, we must check that $t \mapsto \frac{\partial F}{\partial x}(t, W(t)) = e^{W(t)}$ belongs to $\mathscr{M}_T^2$. It is clear that for each t the random variable $e^{W(t)}$ is $\mathscr{F}_t$-measurable. We show next that it belongs to $L^2(\mathbb{P})$ for each t and that the resulting function is square integrable. We have

$$\mathbb{E}\big((e^{W(t)})^2\big) = \mathbb{E}\big(e^{2W(t)}\big) = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} e^{2x} e^{-\frac{x^2}{2t}}\,dx = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} e^{2t} e^{-\frac{(x-2t)^2}{2t}}\,dx = e^{2t}.$$

It follows that $e^{W(t)} \in L^2(\mathbb{P})$, $\|e^{W(t)}\|_{L^2(\mathbb{P})} = e^t$, and therefore $t \mapsto e^{W(t)}$ belongs to $\mathscr{M}_T^2$. Noting that $F(0, W(0)) = F(0,0) = 1$, Itô's formula gives:

$$e^{W(t)} - 1 = F(t, W(t)) - F(0, W(0)) = \int_0^t e^{W(s)}\,dW(s) + \tfrac{1}{2}\int_0^t e^{W(s)}\,ds.$$

Thus, $U(t) = e^{W(t)}$ solves the equation

$$dU(t) = \tfrac{1}{2}U(t)\,dt + U(t)\,dW(t), \qquad U(0) = 1. \tag{3.6}$$

Our guess was too naive! But inspection of the equation that *is* solved by $e^{W(t)}$ leads to a second guess, viz. $U(t) = e^{-\frac{1}{2}t + W(t)}$. In fact, by differentiating formally and using the equation for $e^{W(t)}$ just derived, we obtain

$$\begin{aligned} U'(t) &= \big(e^{-\frac{1}{2}t}\big)' e^{W(t)} + e^{-\frac{1}{2}t}\big(e^{W(t)}\big)' \\ &= -\tfrac{1}{2}e^{-\frac{1}{2}t}e^{W(t)} + e^{-\frac{1}{2}t}\big(\tfrac{1}{2}e^{W(t)} + e^{W(t)}W'(t)\big) = e^{-\frac{1}{2}t}e^{W(t)}W'(t) = U(t)W'(t). \end{aligned}$$

This seems to suggest that our second guess for $U(t)$ is the correct one.

Motivated by these heuristics we take $F(t,x) = e^{-\frac{1}{2}t+x}$ and try to apply Itô's formula. We have

$$\frac{\partial F}{\partial t}(t,x) = -\tfrac{1}{2}e^{-\frac{1}{2}t+x}, \qquad \frac{\partial F}{\partial x}(t,x) = \frac{\partial^2 F}{\partial x^2}(t,x) = e^{-\frac{1}{2}t+x}.$$

From the computations above it follows that $t \mapsto \frac{\partial F}{\partial x}(t, W(t)) = e^{-\frac{1}{2}t+W(t)}$ belongs to $\mathscr{M}_T^2$. Itô's formula gives

$$\begin{aligned} e^{-\frac{1}{2}t+W(t)} - 1 &= -\int_0^t \tfrac{1}{2}e^{-\frac{1}{2}s+W(s)}\,ds + \int_0^t e^{-\frac{1}{2}s+W(s)}\,dW(s) + \tfrac{1}{2}\int_0^t e^{-\frac{1}{2}s+W(s)}\,ds \\ &= \int_0^t e^{-\frac{1}{2}s+W(s)}\,dW(s). \end{aligned}$$

It follows that $U(t) := e^{-\frac{1}{2}t+W(t)}$ is indeed an L^2-solution of the integral equation

$$U(t) = 1 + \int_0^t U(s)\,dW(s), \qquad t \in [0,T].$$

One can also proceed more systematically and try to use an integrating factor to get rid of the additional term $\frac{1}{2}U(t)\,dt$ in (3.6). Accordingly one guesses that $U(t)$ is of the form $\phi(t)e^{W(t)}$ for some $\phi : [0,T] \to \mathbb{R}$. Applying Itô's formula to $F(t,x) = \phi(t)\,e^x$ then leads to a differential equation for ϕ which is solved by $\phi(t) = e^{-\frac{1}{2}t}$.
For arbitrary $\mathscr{F}_0$-measurable $\eta_0 \in L^2(\mathbb{P})$, we deduce that $U(t) := \eta_0 e^{-\frac{1}{2}t+W(t)}$ solves the equation (3.4); cf. the exercise below.

• *Exercise.* Let $\eta_0 \in L^2(\mathbb{P})$ be $\mathscr{F}_0$-measurable, let $U \in \mathscr{M}_T^2$, and assume that η_0 is independent of $U(t)$ for all $t \in [0,T]$. Show that $\eta_0 U \in \mathscr{M}_T^2$ and

$$\int_0^T \eta_0 U(t)\,dW(t) = \eta_0 \int_0^T U(t)\,dW(t).$$

Explain how this was used in the above example.

• *Exercise.* The *Black-Sholes equation*

$$dU(t) = aU(t)\,dt + \sigma U(t)\,dW(t), \qquad U(0) = \eta_0,$$

describes the value $U(t)$ at a stock in a market with interest rate equals a and volatility σ. The initial value of the stock is described by the random variable $\eta_0 \in L^2(\mathbb{P})$ is $\mathscr{F}_0$-measurable (and is usually taken to be constant). Represent this equation as a stochastic integral equation and show that its unique adapted L^2-solution is given by $U(t) = \eta_0 e^{(a-\frac{1}{2}\sigma^2)t+\sigma W(t)}$.

Sketch of the proof of Itô's formula: First we assume that the partial derivatives $\frac{\partial F}{\partial x}$ and $\frac{\partial^2 F}{\partial x^2}$ are bounded.
Of course it is enough to prove the formula for $t = T$. Consider the partition $P_n = \{t_0^n, \ldots, t_n^n\}$ of $[0,T]$ with $t_j^n = \frac{jT}{n}$. As usual we write $\Delta_j^n W = W(t_{j+1}^n) - W(t_j^n)$. By Taylor's formula, for

each $\omega \in \Omega$ there exist $w_j^n(\omega) \in \big(W(t_j^n,\omega), W(t_{j+1}^n,\omega)\big)$ and there exist $s_j^n \in (t_j^n, t_{j+1}^n)$ such that

$$
\begin{aligned}
F(T,&W(T)) - F(0,W(0)) \\
&= \sum_{j=0}^{n-1}\big(F(t_{j+1}^n, W(t_{j+1}^n)) - F(t_j^n, W(t_j^n))\big) \\
&= \sum_{j=0}^{n-1}\big(F(t_{j+1}^n, W(t_{j+1}^n)) - F(t_j^n, W(t_{j+1}^n))\big) + \sum_{j=0}^{n-1}\big(F(t_j^n, W(t_{j+1}^n)) - F(t_j^n, W(t_j^n))\big) \\
&= \sum_{j=0}^{n-1}\frac{\partial F}{\partial t}(s_j^n, W(t_{j+1}^n))\Delta_j^n t + \sum_{j=0}^{n-1}\frac{\partial F}{\partial x}(t_j^n, W(t_j^n))\Delta_j^n W + \tfrac12\sum_{j=0}^{n-1}\frac{\partial^2 F}{\partial x^2}(t_j^n, w_j^n)(\Delta_j^n W)^2 \\
&= \sum_{j=0}^{n-1}\frac{\partial F}{\partial t}(s_j^n, W(t_{j+1}^n))\Delta_j^n t + \sum_{j=0}^{n-1}\frac{\partial F}{\partial x}(t_j^n, W(t_j^n))\Delta_j^n W + \tfrac12\sum_{j=0}^{n-1}\frac{\partial^2 F}{\partial x^2}(t_j^n, W(t_j^n))\Delta_j^n t \\
&\qquad + \tfrac12\sum_{j=0}^{n-1}\frac{\partial^2 F}{\partial x^2}(t_j^n, W(t_j^n))\big((\Delta_j^n W)^2 - \Delta_j^n t\big) \\
&\qquad + \tfrac12\sum_{j=0}^{n-1}\left[\frac{\partial^2 F}{\partial x^2}(t_j^n, w_j^n) - \frac{\partial^2 F}{\partial x^2}(t_j^n, W(t_j^n))\right](\Delta_j^n W)^2.
\end{aligned}
$$

These five sums are considered separately. For the first sum we can show that

$$\lim_{n\to\infty}\sum_{j=0}^{n-1}\frac{\partial F}{\partial t}(s_j^n, W(t_{j+1}^n))\Delta_j^n t = \int_0^b \frac{\partial F}{\partial t}(t, W(t))\,dt$$

a.s., using the continuity of the paths of $\{W(t)\}_{t\geqslant 0}$ and the uniform continuity of F on sets of the form $[0,b]\times[-M,M]$. For the second sum we obtain

$$\lim_{n\to\infty}\sum_{j=0}^{n-1}\frac{\partial F}{\partial x}(t_j^n, W(t_j^n))\Delta_j^n W = \int_0^b \frac{\partial F}{\partial x}(t, W(t))\,dW(t)$$

in $L^2(\mathbb{P})$. For the third sum we have

$$\lim_{n\to\infty}\tfrac12\sum_{j=0}^{n-1}\frac{\partial^2 F}{\partial x^2}(t_j^n, W(t_j^n))\Delta_j^n t = \tfrac12\int_0^t \frac{\partial^2 F}{\partial x^2}(t, W(t))\,dt,$$

again by continuity. For the fourth sum we have

$$\lim_{n\to\infty}\tfrac12\sum_{j=0}^{n-1}\frac{\partial^2 F}{\partial x^2}(t_j^n, W(t_j^n))\big((\Delta_j^n W)^2 - \Delta_j^n t\big) = 0$$

in $L^2(\mathbb{P})$; the main ingredient here is Theorem 2.2. Finally, for the fifth sum we have

$$\lim_{n\to\infty}\tfrac12\sum_{j=0}^{n-1}\left[\frac{\partial^2 F}{\partial x^2}(t_j^n, w_j^n) - \frac{\partial^2 F}{\partial x^2}(t_j^n, W(t_j^n))\right](\Delta_j^n W)^2 = 0$$

a.s., again by continuity.
The boundedness assumption on the partial derivatives of F is removed as follows. For each n we choose a compactly supported smooth function $\phi_n : \mathbb{R} \to [0,1]$ with $\phi_n(x) = 1$ for all $x \in [-n,n]$. Then

$$F_n(t,x) = \phi_n(x)F(t,x)$$

has bounded first and second partial derivatives with respect to x. Therefore by the first part of the proof, Itô's formula holds for each F_n.
Let

$$\Omega_n = \left\{\omega \in \Omega : \sup_{t\in[0,T]} |W(t,\omega)| \leqslant n\right\}.$$

For $\omega \in \Omega_n$ we have

$$F(t, W(t,\omega)) = F_n(t, W(t,\omega)), \qquad t \in [0,T],$$

and therefore Itô's formula holds a.e. on each Ω_n. Since the paths $t \mapsto W(t,\omega)$ are continuous $\mathbb{P}$-a.s., the set $\Omega \setminus \bigcup_{n\geqslant 1} \Omega_n$ is a $\mathbb{P}$-null set.

Chapter 4

Diffusions

4.1 Introduction

Consider again our model problem

$$\begin{aligned} U'(t) &= -aU(t) + \sigma W'(t), \quad t \geqslant 0, \\ U(0) &= u, \end{aligned} \tag{4.1}$$

where $a \geqslant 0$, $\sigma \in \mathbb{R}$, and W' denotes white noise. We have interpreted this equation as the integral equation

$$U(t) = u - a \int_0^t U(s)\, ds + \sigma W(t), \quad t \geqslant 0, \tag{4.2}$$

and found its solution by means of the Wiener integral:

$$U(t, u) = e^{-at} u + \sigma \int_0^t e^{-a(t-s)} dW(s). \tag{4.3}$$

The stochastic process $(U(t,u))_{t \geqslant 0}$ is called the *Ornstein-Uhlenbeck process*. For every t the random variable $\int_0^t e^{-a(t-s)} dW(s)$ is centered Gaussian with variance equal to

$$\| s \mapsto e^{-a(t-s)} \|^2_{L^2[0,t]} = \int_0^t e^{-2a(t-s)} ds = (1 - e^{-2at})/2a.$$

Therefore $U(t,u)$ is Gaussian with mean $e^{-at}u$ and variance $\sigma^2(1 - e^{-2at})/2a$.
We have introduced the equation as a model for the velocity of a small particle in a fluid. It belongs to the class of equations of the form

$$\begin{aligned} X'(t) &= a(t, X(t))dt + \sigma(t, X(t))W'(t), \quad t \geqslant 0, \\ X(0) &= X_0, \end{aligned}$$

which can all be interpreted as models for the motion of a small particle in a fluid. The solutions of such equations are therefore called *diffusions*. The study of diffusions includes various aspects of their evolution in time and there is a remarkable connection with deterministic partial differential equations. We will introduce some of the notions for our special case (4.1).

Equation (4.1) describes an evolution of random variables. For many purposes the actual interest is in their distributions. Can we reformulate the equation as an evolution process for the distributions? We can do so in a weak sense. More precisely, we will consider the functions $(t,u) \mapsto \mathbb{E}f(U(t,u))$, for $f \in C_b(\mathbb{R})$. In contrast to Chapter 1, we let in this chapter $C_b(\mathbb{R})$ denote the space of *real valued* bounded continuous functions on $\mathbb{R}$.

4.2 The transition operators

Define for each $t \geqslant 0$ and $f \in C_b(\mathbb{R})$ the *transition operator* $T(t)$ by

$$(T(t)f)(u) = \mathbb{E}f(U(t,u)), \quad u \in \mathbb{R}.$$

By means of the density of the distribution $\mu_{U(t,u)}$ of $U(t,u)$ with respect to the Lebesgue measure, we can write

$$\begin{aligned} (T(t)f)(u) &= \mathbb{E}f(U(t,u)) \\ &= \int_\Omega f(U(t,u)(\omega))d\mathbb{P}(\omega) \\ &= \int_\mathbb{R} f(x)d\mu_{U(t,u)}(x) \\ &= \int_\mathbb{R} f(x)p(e^{-at}u, \tfrac{\sigma^2}{2a}(1-e^{-2at});x)\,dx, \end{aligned}$$

where

$$p(m,\rho;x) = \tfrac{1}{\sqrt{2\pi\rho}}\, e^{-\frac{1}{2}\frac{(x-m)^2}{\rho}}, \quad m,x \in \mathbb{R}, \rho \geqslant 0.$$

In a weak sense, the operators $(T(t))_{t\geqslant 0}$ describe the evolution of the distributions of $(U(t,u))_{t\geqslant 0}$.
Assume that $\mathbf{a > 0}$.

Lemma 4.1. *Let $t \geqslant 0$.*

(i) For every $f \in C_b(\mathbb{R})$ we have that $T(t)f \in C_b(\mathbb{R})$.

(ii) $T(t)$ is a bounded linear operator on $C_b(\mathbb{R})$ and $\|T(t)\| \leqslant 1$.

Proof. (i): First let $u, u_1, u_2, \ldots \in \mathbb{R}$ be such that $u_n \to u$. Then it is clear from (4.3) that $U(t,u_n) \to U(t,u)$ pointwise on Ω. As f is continuous, we have that $f(U(t,u_n)) \to f(U(t,u))$. Because f is bounded, Lebesgue's dominated convergence theorem yields that

$$\int_\Omega f(U(t,u_n)(\omega))\,d\mathbb{P}(\omega) \to \int_\Omega f(U(t,u)(\omega))\,d\mathbb{P}(\omega),$$

which means that $T(t)f$ is continuous at u.
To show that $T(t)f$ is a bounded function, observe that

$$\begin{aligned} |(T(t)f)(u)| &= = |\mathbb{E}f(U(t,u))| = |\int_\Omega f(U(t,u)(\omega))\,d\mathbb{P}(\omega)| \qquad (4.4) \\ &\leqslant \int_\Omega |f(U(t,u)(\omega))|\,d\mathbb{P}(\omega) \leqslant \int_\Omega \|f\|_\infty d\mathbb{P} \leqslant \|f\|_\infty \end{aligned}$$

for all $u \in \mathbb{R}$.
(ii): By (i) we have that $T(t)$ maps $C_b(\mathbb{R})$ to $C_b(\mathbb{R})$ and it is clear that it is linear. Formula (4.4) yields that $\|T(t)f\|_\infty \leqslant \|f\|_\infty$, which means that $T(t)$ is bounded and $\|T(t)\| \leqslant 1$. □

We will now show that the family $(T(t))_{t\geqslant 0}$ has the *semigroup property*, i.e., $T(t+s) = T(t)T(s)$ for all $s, t \geqslant 0$.

Theorem 4.1. *$(T(t))_{t\geqslant 0}$ is a semigroup of bounded linear operators on $C_b(\mathbb{R})$.*

Proof. In addition to Lemma 4.1 we only have to prove the semigroup property. Denote for $t \geqslant 0$ the distribution of $\int_0^t e^{-a(t-s)}dW(s)$ by μ_t, i.e., μ_t is a Gaussian measure on $\mathbb{R}$ with mean 0 and variance $(1-e^{-2at})/2a$. For $t \geqslant 0$ and $u \in \mathbb{R}$ we have that

$$\begin{aligned}(T(t)f)(u) &= \mathbb{E}f(U(t,u))\\ &= \mathbb{E}f\left(e^{-at}u + \sigma\int_0^t e^{-a(t-\tau)}dW(\tau)\right)\\ &= \int_{\mathbb{R}} f\left(e^{-at}u + \sigma x\right)d\mu_t(x).\end{aligned}$$

Let $s, t \geqslant 0$, let $f \in C_b(\mathbb{R})$, and let $u \in \mathbb{R}$. By using the above formula twice, we obtain

$$\begin{aligned}(T(t)T(s)f)(u) &= \int_{\mathbb{R}} (T(s)f)(e^{-at}u + \sigma x)d\mu_t(x)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f\left(e^{-as}(e^{-at}u + \sigma x) + \sigma y\right) d\mu_s(y)d\mu_t(x)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f\left(e^{-a(t+s)}u + \sigma e^{-as}x + \sigma y\right) d\mu_s(y)d\mu_t(x).\end{aligned}$$

To write out $(T(t+s)f)(u)$, observe that

$$\begin{aligned}U(s+t,u) &= e^{-a(s+t)}u + \sigma\int_0^{t+s} e^{-a(t+s-\tau)}dW(\tau)\\ &= e^{-a(s+t)}u + \sigma\int_0^{t} e^{-a(t+s-\tau)}dW(\tau) + \sigma\int_t^{t+s} e^{-a(t+s-\tau)}dW(\tau)\\ &= e^{-a(s+t)}u + \sigma e^{-as}\int_0^{t} e^{-a(t-\tau)}dW(\tau) + \sigma\int_t^{t+s} e^{-a(t+s-\tau)}dW(\tau).\end{aligned}$$

Denote, for abbreviation,

$$\begin{aligned}X &= \int_0^t e^{-a(t-\tau)}dW(\tau), \text{ and}\\ Y &= \int_t^{t+s} e^{-a(t+s-\tau)}dW(\tau) = \int_0^{t+s} \mathbf{1}_{[t,t+s]}(\tau)e^{-a(t+s-\tau)}dW(\tau).\end{aligned}$$

We see that Y is a centered Gaussian random variable with

$$\operatorname{var} Y = \|\mathbf{1}_{[t,t+s]}(\cdot)e^{-a(t+s-\cdot)}\|^2_{L^2[0,t+s]} = \int_t^{t+s} e^{-2a(t+s-\tau)}d\tau = \frac{1-e^{-2as}}{2a}.$$

Hence the distributions of X and Y are the measures μ_t and μ_s, respectively. Moreover, we have that X and Y are independent, because they are both centered Gaussians and

$$\begin{aligned}\mathbb{E}XY &= \mathbb{E}\left(\int_0^t e^{-a(t-\tau)}dW(\tau)\right)\left(\int_t^{t+s} e^{-a(t+s-\tau)}dW(\tau)\right)\\ &= \mathbb{E}\left(\int_0^{t+s} \mathbf{1}_{[0,t]}(\tau)e^{-a(t-\tau)}dW(\tau)\right)\left(\int_0^{t+s} \mathbf{1}_{[t,t+s]}(\tau)e^{-a(t+s-\tau)}dW(\tau)\right)\\ &= \int_0^{t+s} \mathbf{1}_{[0,t]}(\tau)e^{-a(t-\tau)}\mathbf{1}_{[t,t+s]}(\tau)e^{-a(t+s-\tau)}d\tau = 0\end{aligned}$$

It follows that the distribution of the random vector (X,Y) is the product measure $\mu_t \otimes \mu_s$. Now we find that

$$\begin{aligned}(T(t+s)f)(u) &= \mathbb{E}f(U(t+s,u))\\ &= \mathbb{E}f\left(e^{-a(t+s)}u + \sigma e^{-as}X + \sigma Y\right)\\ &= \int_{\mathbb{R}^2} f\left(e^{-a(t+s)}u + \sigma e^{-as}x + \sigma y\right)\, d(\mu_t \otimes \mu_s)(x,y)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f\left(e^{-a(t+s)}u + \sigma e^{-as}x + \sigma y\right)\, d\mu_t(x)d\mu_s(y),\end{aligned}$$

so that $(T(t+s)f)(u) = (T(t)T(s)f)(u)$. □

Because of the above theorem, the family of transition operators $(T(t))_{t\geqslant 0}$ is called the *transition semigroup* of equation (4.1).

4.3 The invariant measure

What can we say about the distributions of $U(t,u)$ for $t \to \infty$?

Proposition 4.2. *For every $u \in \mathbb{R}$ the distributions of the random variables $U(t,u)$ converge as $t \to \infty$ to a Gaussian measure with mean 0 and variance $\sigma^2/2a$.*

Proof. The distribution of $U(t,u)$ is a Gaussian measure and

$$\mathbb{E}U(t,u) = e^{-at}u \to 0$$

and

$$\operatorname{var} U(t,u) = \sigma^2(1-e^{-2at})/2a \to \sigma^2/2a \quad \text{as } t \to \infty.$$

Then Proposition 1.24 yields the assertion. □

The limiting distribution has the property that it is ‘stationary’. To define this notion we need to consider stochastic initial conditions.

As $U(t,u)$ is a random variable, we can as well consider an initial condition that is a random variable. Let U_0 be a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ and consider

$$U(t) = U_0 - a\int_0^t U(s)ds + \sigma W(t), \quad t \geqslant 0. \tag{4.5}$$

Exercise 4.3. Assume that $U_0 \in L^2(\mathbb{P})$. Check that

$$U(t, U_0) = e^{-at}U_0 + \sigma \int_0^t e^{-a(t-s)} dW(s), \quad t \geqslant 0, \tag{4.6}$$

is a solution of (4.5).

Often initial conditions are considered that are 'independent of $(W(t))_{t\geqslant 0}$'. A random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ is called *independent* of $(W(t))_{t\geqslant 0}$ if the σ-algebra that it generates is independent of the one generated by all $W(t)$, $t \geqslant 0$. For instance, if $(\mathscr{F}_t)_{t\geqslant 0}$ is a filtration in $\mathscr{F}$ and $(W(t))_{t\geqslant 0}$ is adapted to it as a Brownian motion, then every random variable that is $\mathscr{F}_0$ measurable is independent of $(W(t))_{t\geqslant 0}$.

Lemma 4.2. *If U_0 is a random variable that is independent of $(W(t))_{t\geqslant 0}$, then U_0 is independent of $\int_0^t e^{-a(t-s)} dW(s)$ for every $t \geqslant 0$.*

Proof. The assumption means that the σ-algebra generated by U_0 and the σ-algebra $\mathscr{A}$ generated by $\{W(t)\}_{t\geqslant 0}$ are independent. By construction, $\int_0^t e^{-a(t-s)} dW(s)$ is measurable with respect to $\mathscr{A}$, so that U_0 and $\int_0^t e^{-a(t-s)} dW(s)$ are independent for all $t \geqslant 0$. $\square$

Note that there may not exist a non-trivial random variable on Ω that is independent of $(W(t))_{t\geqslant 0}$. However, by the product construction of Section 1.2 one can always extend the space to generate independent random variables.
If U_0 is a random variable in $L^2(\mathbb{P})$ that is independent of $(W(t))_{t\geqslant 0}$ and such that the distribution of $U(t, U_0)$ equals the distribution of U_0 for every $t \geqslant 0$, then that distribution is called *stationary*.

Theorem 4.4. *The Gaussian distribution with mean 0 and variance $\sigma^2/2a$ is a stationary distribution for (4.5). That means: if U_0 is a Gaussian random variable with mean 0 and variance $\sigma^2/2a$ that is independent of $(W(t))_{t\geqslant 0}$, then $U(t, U_0)$ is centered Gaussian with variance $\sigma^2/2a$ for every $t \geqslant 0$.*

Proof. Being the sum of two independent Gaussians, $U(t, U_0)$ is Gaussian. Moreover,

$$\mathbb{E}U(t, U_0) = \mathbb{E}e^{-at}U_0 + \mathbb{E}\sigma \int_0^t e^{-a(t-s)} dW(s) = 0$$

and

$$\begin{aligned} \operatorname{var} U(t, U_0) &= \mathbb{E}\left(e^{-at}U_0\right)^2 + \mathbb{E}\left(\sigma \int_0^t e^{-a(t-s)} dW(s)\right)^2 \\ &= e^{-2at}\sigma^2/2a + \sigma^2(1 - e^{-2at})/2a \\ &= \sigma^2/2a, \end{aligned}$$

for all $t \geqslant 0$. $\square$

Theorem 4.5. *The Gaussian distribution with mean 0 and variance $\sigma^2/2a$ is the only stationary distribution for (4.5).*

Proof. We will show that for an arbitrary initial condition $U_0 \in L^2(\mathbb{P})$ that is independent of $(W(t))_{t\geqslant 0}$, the distributions $\mu_{X(t)}$ of $U(t)$ converge to the Gaussian measure μ_∞ with mean 0 and variance $\sigma^2/2a$. Then it follows that the distribution of U_0 is stationary only if it equals μ_∞.
Let $U_0 \in L^2(\mathbb{P})$ be independent of $(W(t))_{t\geqslant 0}$ and let μ_0 be its distribution. Then U_0 and $\sigma^2 \int_0^t e^{-a(t-s)}dW(s)$ are independent for all $t \geqslant 0$, so that (4.6) yields that for $\xi \in \mathbb{R}$

$$\begin{aligned}
\hat{\mu}_{U(t,U_0)}(\xi) &= \mathbb{E}e^{-i\xi U(t,U_0)} \\
&= \int_{\mathbb{R}^2} e^{-i\xi e^{-at}x} e^{-i\xi y} d(\mu_0 \otimes \mu_{\sigma^2 \int_0^t e^{-a(t-s)}dW(s)})(x,y) \\
&= \int_{\mathbb{R}} e^{-i\xi e^{-at}x} d\mu_0(x) \int_{\mathbb{R}} e^{-i\xi y} d\mu_{\sigma^2 \int_0^t e^{-a(t-s)}dW(s)}(y) \\
&= \int_{\mathbb{R}} e^{-i\xi e^{-at}x} d\mu_0(x) \hat{\mu}_{\sigma^2 \int_0^t e^{-a(t-s)}dW(s)}(\xi).
\end{aligned}$$

The first factor converges as $t \to \infty$, by the Lebesgue dominated convergence theorem, to $\int_{\mathbb{R}} 1 d\mu_0(x) = 1$ and the second, by Proposition 4.2, to the Fourier transform of μ_∞ at ξ. Hence, $\hat{\mu}_{U(t,U_0)}(\xi) \to \hat{\mu}_\infty(\xi)$ for every $\xi \in \mathbb{R}$, so that $\mu_{U(t,U_0)}$ converges to μ_∞ as $t \to \infty$. $\square$

We can formulate stationarity by means of the transition operators introduced in the previous section.

Definition 4.6. A Borel probability measure μ on $\mathbb{R}$ is called an *invariant measure* for (4.5) if

$$\int_{\mathbb{R}} (T(t)f)(u)\, d\mu(u) = \int_{\mathbb{R}} f(u)\, d\mu(u) \quad \text{for all } f \in C_b(\mathbb{R}) \text{ and } t \geqslant 0,$$

or, explicitly,

$$\int_{\mathbb{R}} \mathbb{E}f(U(t,u))d\mu(u) = \int_{\mathbb{R}} f(u)d\mu \quad \text{for all } f \in C_b(\mathbb{R}) \text{ and } t \geqslant 0.$$

Proposition 4.7. *Let U_0 be a random variable that is independent of $(W(t))_{t\geqslant 0}$ and let μ_0 be its distribution. Then the distribution of U_0 is stationary if and only if μ_0 is an invariant measure.*

Proof. Let $t \geqslant 0$. Denote $V = \int_0^t e^{-a(t-s)}dW(s)$ and let μ_V be its distribution. Since U_0 and V are independent, their joint distribution is the product measure $\mu_0 \otimes \mu_V$. Therefore

$$\begin{aligned}
\mathbb{E}f(U(t,U_0)) &= \mathbb{E}f\left(e^{-at}U_0 + \sigma \int_0^t e^{-a(t-s)}dW(s)\right) \\
&= \int_{\mathbb{R}^2} f(e^{-at}u + \sigma v) d(\mu_0 \otimes \mu_V)(u,v) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(e^{-at}u + \sigma v) d\mu_0(u) d\mu_V(v) \\
&= \int_{\mathbb{R}} \mathbb{E}f\left(e^{-at}u + \sigma \int_0^t e^{-a(t-s)}dW(s)\right) d\mu_0(u) \\
&= \int_{\mathbb{R}} \mathbb{E}f(U(t,u)) d\mu_0(u) \\
&= \int_{\mathbb{R}} (T(t)f)(u) d\mu_0(u).
\end{aligned}$$

If μ_0 is an invariant measure, i.e.,

$$\int_{\mathbb{R}} (T(t)f)(u)d\mu_0(u) = \int_{\mathbb{R}} f(u)d\mu_0(u),$$

then it follows that

$$\mathbb{E}f(U(t,U_0)) = \int_{\mathbb{R}} f(u)d\mu_0(u) = \mathbb{E}f(U_0), \quad \text{for all } f \in C_b(\mathbb{R}),$$

so that the distributions of $U(t,U_0)$ and U_0 are equal.
Conversely, if the distributions of $U(t,U_0)$ and U_0 are equal, then it follows that

$$\begin{aligned}\int_{\mathbb{R}} (T(t)f)(u)d\mu_0(u) &= \mathbb{E}f(U(t,U_0)) \\ &= \mathbb{E}f(U_0) = \int_{\mathbb{R}} f(u)d\mu_0(u), \quad \text{for all } f \in C_b(\mathbb{R}),\end{aligned}$$

which means that μ_0 is an invariant measure. □

It follows that the Gaussian measure on $\mathbb{R}$ with mean 0 and variance $\sigma^2/2a$ is an invariant measure for (4.5).

4.4 The Kolmogorov equation

Let $f \in C_b(\mathbb{R})$ and define

$$\begin{aligned} z(t,x) &= \mathbb{E}f(U(t,x)) = (T(t)f)(x) \\ &= \int_{\mathbb{R}} f(y)p(e^{-at}u, \tfrac{\sigma^2}{2a}(1-e^{-2at}); y)dy \end{aligned}$$

and

$$p(e^{-at}u, \tfrac{\sigma^2}{2a}(1-e^{-2at}); y) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2a}(1-e^{-2at})}} e^{-\frac{1}{2}\frac{(y-e^{-at}u)^2}{\frac{\sigma^2}{2a}(1-e^{-2at})}}.$$

If $\xi \in \mathbb{R}$ and $f(x) = e^{-ix\xi}$, $x \in \mathbb{R}$, then —by considering the real and imaginary parts separately— we can consider the corresponding function z:

$$z(t,x) = \mathbb{E}e^{-iU(t,x)\xi} = e^{-ie^{-at}x\xi - \frac{1}{2}\frac{\sigma^2}{2a}(1-e^{-2at})\xi^2}, \quad t \geqslant 0, \quad x \in \mathbb{R}. \tag{4.7}$$

Differentiation of z yields:

$$\begin{aligned} z_t &= \left\{iaxe^{-at}\xi - \tfrac{1}{2}\sigma^2 e^{-2at}\xi^2\right\} z, \\ z_x &= -ie^{-at}\xi z, \\ z_{xx} &= -e^{-2at}\xi^2 z. \end{aligned}$$

Further, $z(0,x) = e^{-ix\xi}$, $x \in \mathbb{R}$. Hence we find that

$$\begin{cases} z_t(t,x) = \tfrac{1}{2}\sigma^2 z_{xx}(t,x) - axz_x(t,x), & t \geqslant 0,\ x \in \mathbb{R}, \\ z(0,x) = e^{-ix\xi}, & x \in \mathbb{R}. \end{cases}$$

This means that for $f = x \mapsto e^{-ix\xi}$ the corresponding function z satisfies

$$\begin{cases} z_t &= \frac{1}{2}\sigma^2 z_{xx} - axz_x, \quad t > 0,\ x \in \mathbb{R}, \\ z(0,x) &= f(x), \qquad\qquad\qquad x \in \mathbb{R}. \end{cases} \tag{4.8}$$

These equations are called the *Kolmogorov equations* associated with (4.5). By direct verification one can show that the function z given by (4.7) satisfies (4.8) for any $f \in C_b^2(\mathbb{R})$. Define

$$\mathscr{Z} := \{z \in C((0,\infty) \times \mathbb{R}) \colon\ z_t, z_x, z_{xx} \text{ exist and are continuous on } (0,\infty) \times \mathbb{R}\}.$$

Theorem 4.8. *Let $f \in C_b^2(\mathbb{R})$. Then for the function z defined by*

$$z(t,x) = \mathbb{E}f(U(t,x)), \quad t \geqslant 0, x \in \mathbb{R},$$

we have that $z \in \mathscr{Z}$ and

$$\begin{cases} z_t &= \frac{1}{2}\sigma^2 z_{xx} - axz_x, \quad t > 0,\ x \in \mathbb{R}, \\ z(0,x) &= f(x), \qquad\qquad\qquad x \in \mathbb{R}. \end{cases}$$

If $\sigma = 0$, then problem (4.8) is a first order partial differential equation of hyperbolic type. The corresponding stochastic differential equation (4.1) is then in fact deterministic and for all $\omega \in \Omega$, $U(t,x)(\omega)$ satisfies

$$\begin{cases} \frac{dU}{dt}(t) &= -aU(t), \quad t \geqslant 0, \\ U(0) &= x. \end{cases}$$

So in this case, the above considerations reduce to the conclusion that $z(t,x) = \mathbb{E}f(U(t,x)) = f(U(t,x))$ is a solution to (4.8), which is actually the method of characteristics. If $\sigma \neq 0$, (4.8) is a second order parabolic equation and we can use the genuinely stochastic equation (4.1) to find a solution $z(t,x) = \mathbb{E}f(U(t,x))$. This approach may be interpreted as a stochastic method of characteristics.

4.5 The Generator

The differential operator appearing at the right hand side of the Kolmogorov equation is called the 'generator' of the diffusion $(U(t,\cdot))_{t\geqslant 0}$. It is related to the transition operators in the following way.

Definition 4.9. Define for $x \in \mathbb{R}$

$$(Af)(x) = \lim_{t\downarrow 0} \frac{(T(t)f)(x) - f(x)}{t} = \lim_{t\downarrow 0} \frac{\mathbb{E}f(U(t,x)) - f(x)}{t},$$

and let $\mathscr{D}_A$ be the set of functions f in $C_b(\mathbb{R})$ for which the limit exists for all $x \in \mathbb{R}$ and is a member of $C_b(\mathbb{R})$. A is called the *generator* of $(U(t,\cdot))_{t\geqslant 0}$.

Theorem 4.10. *The generator of the diffusion $(U(t,\cdot))_{t\geqslant 0}$ given by (4.1) is*

$$\frac{1}{2}\frac{d^2}{dx^2} - ax\frac{d}{dx}.$$

Proof. For $f \in C_b^2(\mathbb{R})$ the function $z(t,x) = (T(t)f)(x)$, $t \geqslant 0$, $x \in \mathbb{R}$, satisfies the Kolmogorov equation

$$\frac{\partial z}{\partial t} = \frac{1}{2}\sigma^2 \frac{\partial^2 z}{\partial x^2} - ax\frac{\partial z}{\partial x}, \quad x \in \mathbb{R},$$

so that we have for the generator A:

$$\begin{aligned}(Af)(x) &= \frac{\partial z}{\partial t}(0,x) = \tfrac{1}{2}\sigma^2 \frac{\partial^2 z}{\partial x^2}(0,x) - ax\frac{\partial z}{\partial x}(0,x) \\ &= \tfrac{1}{2}\sigma^2 f''(x) - axf'(x), \quad x \in \mathbb{R},\end{aligned}$$

because the function $x \mapsto z(0,x)$ equals f. □

Remark 4.11. With aid of the Itô formula it can be shown that the generator of the stochastic differential equation

$$U'(t) = a(U(t)) + \sigma(U(t))W'(t), \quad t \geqslant 0,$$

is given by

$$(Af)(x) = a(x)f'(x) + \tfrac{1}{2}\sigma(x)^2 f''(x), \quad x \in \mathbb{R},\ f \in C_b^2(\mathbb{R}).$$

Exercise 4.12. Denote the generator of (4.1) by

$$\mathscr{L} = \frac{1}{2}\sigma^2 \frac{d^2}{dx^2} - ax\frac{d}{dx}.$$

Let $f \in C_b^2(\mathbb{R})$ and $g \in C_c^1(\mathbb{R})$ (i.e. g is a C^1-function with compact support). Let μ_∞ be the invariant measure of Theorem 4.4: $\mu_\infty = N(0, \sigma^2/2a)$. Show the following *integration by parts* formula:

$$\tfrac{1}{2}\sigma^2 \int_{\mathbb{R}} f'g'\,d\mu_\infty = -\int_{\mathbb{R}} \mathscr{L}(f)g\,d\mu_\infty.$$

(Hint: let p denote the density of μ_∞ with respect to the Lebesgue measure. Verify that

$$-\tfrac{1}{2}\sigma^2 p'(x) - axp(x) = 0, \quad x \in \mathbb{R},$$

and use ordinary integration by parts.)

It follows that with aid of the invariant measure we can find a weak formulation of the Kolmogorov equation:

$$\int_{\mathbb{R}} z_t(t,x)g(x)p(x)\,dx + \tfrac{1}{2}\sigma^2 \int_{\mathbb{R}} z_x(t,x)g'(x)p(x)\,dx = 0, \quad \forall g \in C_c^1(\mathbb{R}).$$

Moreover, if $f, g \in C_b^2(\mathbb{R}) \cap C_c^1(\mathbb{R})$, we have

$$\int_{\mathbb{R}} \mathscr{L}(f)gp\,dx = -\tfrac{1}{2}\sigma^2 \int_{\mathbb{R}} f'g'p\,dx = \int_{\mathbb{R}} f\mathscr{L}(g)p\,dx,$$

which means that the invariant measure 'symmetrizes' the differential operator $\mathscr{L}$. This observation can be used to prove uniqueness of solutions of (4.8) in a certain class. Exploitation of the above ideas together yields a 'probabilistic method' to solve and prove properties of deterministic partial differential equations.

Finally we will show that the transition semigroup $(T(t))_{t\geqslant 0}$ can be extended to a semigroup on $L^2(\mu_\infty)$ that is *strongly continuous*, i.e., $t \mapsto T(t)f$ is continuous from $[0,\infty)$ to $L^2(\mathbb{P})$ for all $f \in L^2(\mathbb{P})$. Here μ_∞ denotes again the invariant measure. We begin by showing that for every $t \geqslant 0$ the operator $T(t)$ extends uniquely to a bounded linear operator on $L^2(\mu_\infty)$. Indeed, first observe that by Cauchy-Schwarz:

$$(\mathbb{E}X)^2 \leqslant \left(\int_\Omega 1X\, d\mathbb{P}\right)^2 \leqslant \int_\Omega 1^2 d\mathbb{P} \int_\Omega X^2 d\mathbb{P} = \mathbb{E}X^2$$

for any $X \in L^2(\mathbb{P})$. Hence for $f \in C_b(\mathbb{R})$:

$$\begin{aligned}\|(T(t)f)\|^2_{L^2(\mu_\infty)} &= \int_{\mathbb{R}} ((T(t)f)(u))^2\, d\mu_\infty(u) \\ &= \int_{\mathbb{R}} (\mathbb{E}f(U(t,u)))^2\, d\mu_\infty(u) \\ &\leqslant \int_{\mathbb{R}} \mathbb{E}f(U(t,u))^2 d\mu_\infty(u).\end{aligned}$$

Denote $g(x) = f(x)^2$, $x \in \mathbb{R}$. Then

$$\begin{aligned}\|T(t)f\|^2_{L^2(\mu_\infty)} &\leqslant \int_{\mathbb{R}} \mathbb{E}g(U(t,u))d\mu_\infty(u) \\ &= \int_{\mathbb{R}} g(u)d\mu_\infty(u) = \|f\|^2_{L^2(\mu_\infty)}.\end{aligned}$$

Hence $T(t)$ extends uniquely to a bounded linear operator on $L^2(\mu_\infty)$ that we will again denote by $T(t)$, and $\|T(t)\| \leqslant 1$.

Theorem 4.13. *$(T(t))_{t\geqslant 0}$ is a strongly continuous semigroup on $L^2(\mathbb{P})$.*

Proof. Let $u \in \mathbb{R}$. As we have seen in Chapter 2, we can choose a version $(\tilde{U}(t,u))_{t\geqslant 0}$ of $(U(t,u))_{t\geqslant 0}$ with continuous paths, i.e., for every t there is a $\tilde{U}(t,u)$ in the equivalence class of the $L^2(\mathbb{P})$-element $U(t,u)$ such that $t \mapsto \tilde{U}(t,u)(\omega)$ is continuous for $\mathbb{P}$-almost every $\omega \in \Omega$.
Let $f \in C_b(\mathbb{R})$, let $t \geqslant 0$ and let $(t_n)_n$ be a sequence in $[0,\infty)$ such that $t_n \to t$. Then as $n \to \infty$ we have that

$$\tilde{U}(t_n,u)(\omega) \to \tilde{U}(t,u)(\omega) \text{ for a.e. } \omega$$

and then

$$f(\tilde{U}(t_n,u)(\omega)) \to f(\tilde{U}(t,u)(\omega)) \text{ for a.e. } \omega.$$

Since f is bounded, Lebesgue's dominated convergence theorem yields that

$$\mathbb{E}f(\tilde{U}(t_n,u)) \to \mathbb{E}f(\tilde{U}(t,u)),$$

hence

$$\mathbb{E}f(U(t_n,u)) \to \mathbb{E}f(U(t,u)).$$

By dominated convergence once more, we find that

$$\int_{\mathbb{R}} \Big(\mathbb{E}f(U(t_n,u)) - \mathbb{E}f(U(t,u))\Big)^2 d\mu_\infty(u) \to 0,$$

which means that

$$\int_{\mathbb{R}} (T(t_n)f - T(t)f)^2 d\mu_\infty \to 0 \text{ as } n \to \infty.$$

Thus, $t \mapsto T(t)f$ is continuous from $[0, \infty)$ to $L^2(\mathbb{P})$ for every $f \in C_b(\mathbb{R})$.
For an $f \in L^2(\mu_\infty)$, there is for every $\varepsilon > 0$ a $g \in C_b(\mathbb{R})$ with $\|f - g\|_{L^2(\mu_\infty)} < \varepsilon$. Then (because $\|T(t)\| \leqslant 1$):

$$\begin{aligned}\|T(t_n)f - T(t)f\|_{L^2(\mu_\infty)} &\leqslant \|T(t_n)f - T(t_n)g\| + \|T(t_n)g - T(t)g\| + \|T(t)g - T(t)f\| \\ &\leqslant \|f - g\| + \|T(t_n)g - T(t)g\| + \|g - f\| < 2\varepsilon\end{aligned}$$

for $|t_n - t|$ small, which yields the strong continuity on all of $L^2(\mu_\infty)$. □

Exercise 4.14. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with two Brownian motions $(W_1(t))_{t\geqslant 0}$ and $(W_2(t))_{t\geqslant 0}$ that are independent, i.e., the σ-algebra generated by $\{W_1(t)\}_{t\geqslant 0}$ is independent of the σ-algebra generated by $\{W_2(t)\}_{t\geqslant 0}$. Consider the two dimensional diffusion $U(t, x_1, x_2) = (x_1 + W_1(t), x_2 + W_2(t))$, $t \geqslant 0$, where $x_1, x_2 \in \mathbb{R}$.

1. Let $f \in C_b^2(\mathbb{R}^2)$ and define

$$z(t, x_1, x_2) := \mathbb{E}f(U(t, x_1, x_2)), \quad t \geqslant 0,\ x_1, x_2 \in \mathbb{R}.$$

   Show that z satisfies

$$\frac{\partial z}{\partial t} = \tfrac{1}{2}\Delta z,$$

   where

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}.$$

   (Hint: compute $\mathbb{E}f(U(t, x_1, x_2))$ with aid of the knowledge that $U(t, x_1, x_2)$ is a Gaussian random vector with mean (x_1, x_2) and covariance matrix $\begin{pmatrix} t & 0 \\ 0 & t \end{pmatrix}$.)

2. The (2-dimensional) generator of $(U(t, \cdot, \cdot))_{t\geqslant 0}$ is defined by

$$(Af)(x_1, x_2) = \lim_{t\downarrow 0} \frac{\mathbb{E}f(U(t, x_1, x_2)) - f(x_1, x_2)}{t},$$

   where the functions $f\colon \mathbb{R}^2 \to \mathbb{R}$ are twice continuously differentiable. Conclude that

$$A = \tfrac{1}{2}\Delta.$$

Exercise 4.15. Consider the diffusion $U(t, x_1, x_2) = (x_1 + t, x_2 + W(t))$, $t \geqslant 0$, $x_1, x_2 \in \mathbb{R}$. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be twice continuously differentiable. Show that $z\colon (t, x_1, x_2) \mapsto \mathbb{E}f(U(t, x_1, x_2))$ satisfies

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x_1} + \frac{1}{2}\frac{\partial^2 z}{\partial x_2^2}.$$

Bibliography

[1] Arnold, Ludwig, *Stochastic differential equations: theory and applications*, 1992

[2] Breiman, Leo, *Probability,* Addison-Wesley, Reading, Mass., 1968.

[3] Brzeźniak, Z. and T. Zastawniak, *Basic Stochastic Processes*, Springer Verlag, 1999.

[4] Chung, K.L. and R.J. Williams, *Introduction to Stochastic Integration*, 2nd ed., Birkhauser Verlag, 1990.

[5] Clément, Ph. and O.W. van Gaans, An analytical introduction to stochastic differential equations: Part I — The Langevin Equation, *Rend. Istit. Mat. Univ. Trieste* **32** (2000), 1–51.

[6] Da Prato, Giuseppe, *Introduction to differential stochastic equations,* 2nd ed., Appunti, Scuola Normale Superiore, Pisa, 1998.

[7] Da Prato, G. and Zabczyk, J., *Stochastic equations in infinite dimensions* (Encyclopedia of mathematics and its applications 45), Cambridge University Press, 1992.

[8] Halmos, Paul R., *Measure Theory,* Van Nostrand, New York, 1950.

[9] Malliavin, Paul, *Integration and Probability,* (GTM 157) (transl. from the French), Springer-Verlag, New York, 1995.

[10] Oksendal, Bernt, *Stochastic Differential Equations*, 5th edition, Springer Verlag, Berlin - Heidelberg, 1998.

[11] Shiryayev, A.N., *Probability,* (transl. from the Russian) 2nd ed., Springer, New York, 1996.

[12] Stromberg, Karl R., *Probability for Analysts,* Chapman and Hall, New York, 1994.

[13] Stroock, Daniel W. and Varadhan, S.R. Srinivasa, *Multidimensional Diffusion Processes,* Springer, Berlin, 1979.