STOCHASTIC ANALYSIS

F. den Hollander
H. Maassen

Mathematical Institute
University of Nijmegen
Toernooiveld 1
6525 ED Nijmegen
The Netherlands

august 2000

References:

1. F. Black and M. Scholes, The pricing of options and corporate liabilities, *J. Polit. Econom.* **81**, 637-659, 1973.
2. R. Cameron, W.T. Martin, Transformations of Wiener integrals under translations, *Ann. Math. 2* **45**, 386–396, 1944.
3. R.A. Carmona, S.A. Molchanov, *Parabolic Anderson problem and intermittency*, AMS Memoir 518, American Mathematical Society, Providence RI, 1994.
4. K.L. Chung and R. Williams, *Introduction to Stochastic Integration*, Birkhäuser, Boston, 1990 (2nd edition).
5. E.B. Davies, *One-parameter semigroups*, Academic Press 1980.
6. R. Durrett, *Stochastic calculus. A practical introduction*, Probability and Stochastic Series. CRC Press 1996.
7. E.B.Dynkin, The optimum choice of the instant for stopping a Markov process, *Soviet Mathematics* **4**, 627–627, 1963.
8. I.V. Girsanov, On transforming a certain class of stochastic processes by absolutely continuous substitution of measures, *Theory of Probability and Applications* **5**, 285–301, 1960.
9. P.R. Halmos, *Measure Theory*, van Nostrand Reinhold, New York, 1950.
10. R. van der Hofstad, F. den Hollander, W. König, Central limit theorem for the Edwards model, *Ann. Probab.* **25**, 573–597, 1997.
11. K. Itô, *Lectures on Stochastic Processes*, Tata Institute of Fundamental Research, Bombay, 1961.
12. G. Kallianpur, *Stochastic Filtering Theory*, Springer, New York, 1980.
13. R.E. Kalman and R.S. Bucy, New results in linear filtering and prediction theory, Trans. ASME Ser. DJ. Basic Eng. 83, 95-108, 1961.
14. T. Mikosh, *Elementary stochastic calculus with finance in view*, Advanced Series on Statistical Science & Applied Probability **6**, 1999.
15. B. Øksendal, *Stochastic Differential Equations: an introduction with applications*, Springer Berlin, Heidelberg. 1-st edition 1985, 5-th edition 1998.
16. L.S. Ornstein and G.E. Uhlenbeck, On the theory of Brownian motion, Phys. Rev. 36, 823-841, 1930.
17. N. Wiener, Differential space, *J. Math. and Phys.* **2**, 131–174, 1923.
18. N. Wiener, The homogeneous chaos, *Amer. J. Math.* **60**, 897–936, 1938.
19. D. Williams, *Probability with martingales*, Cambridge University Press, Cambridge 1991.

The notes in this syllabus are based primarily on the book by Øksendal.

Contents

1 Introduction

In stochastic analysis one studies random functions of one variable and various kinds of integrals and derivatives thereof. The argument of these functions is usually interpreted as "time", so that the functions themselves can be thought of as paths of random processes.

Here, like in other areas of mathematics, going from the discrete to the continuous yields a pay-off in simplicity and smoothness, at the price of more complicated analysis. Compare, to make an analogy, the integral $\int_0^t x^3 dx$ with the sum $\sum_{k=1}^n k^3$. The integral requires a more refined analysis for its definition and for the proof of its properties, but once this has been done the integral is easier to calculate. Similarly, in stochastic analysis you will become acquainted with a convenient differential calculus as a reward for some hard work in analysis.

1.1 Applications

Stochastic analysis can be applied in a wide variety of situations. We sketch a few examples below.

1. Some differential equations become more realistic when we allow some randomness in their coefficients. Consider for example the following *growth equation*, used among other areas in population biology:

$$\frac{d}{dt}S_t = (r + \text{“}N_t\text{”})S_t. \tag{1.1}$$

   Here, S_t is the size of the population at time t, r is the average growth rate of the population, and the "noise" N_t models random fluctuations in the growth rate.

2. The *Langevin equation* describes the behaviour of a dust particle suspended in a liquid:

$$m\frac{d}{dt}V_t = -\eta V_t + \text{“}N_t\text{”}. \tag{1.2}$$

   Here, V_t is the velocity at time t of the dust particle, the friction exerted on the particle due to the viscosity η of the liquid is $-\eta V_t$, and the "noise" N_t stands for the disturbance due to the thermal motion of the surrounding liquid molecules colliding with the particle. This equation is fundamental in physics.

3. The path of the dust particle in example 2 is observed with some inaccuracy. One measures the perturbed signal Z_t given by

$$Z_t = V_t + \text{“}\tilde{N}_t\text{”}. \tag{1.3}$$

   Here $\tilde{N}_t$ is again a "noise". One is interested in the best guess for the actual value of V_t, given the observation Z_s for $0 \le s \le t$. This is called a *filtering problem*: how to filter away the noise $\tilde{N}_t$. Kalman and Bucy (1961) found a linear algorithm, which was almost immediately applied in aerospace engineering. Filtering theory is now a flourishing and extremely useful discipline.

4. Stochastic analysis can help solve boundary value problems such as the *Dirichlet problem*. If the value of a harmonic function f on the boundary of some bounded regular region $D \subset \mathbb{R}^n$ $(n \ge 1)$ is known, then the value of f in the interior of D can be expressed as follows:

$$\mathbb{E}\left(f\left(B_\tau^x\right)\right) = f(x), \tag{1.4}$$

where "$B_t^x := x + \int_0^t N_s ds$" is an "integrated noise" or *Brownian motion*, starting at x, and τ denotes the time when this Brownian motion first reaches the boundary. (A harmonic function f is a function satisfying $\Delta f = 0$ with Δ the Laplacian.)

5. Stochastic analysis has found extensive application nowadays in finance. A typical problem is the following. At time $t = 0$ an investor buys stocks and bonds on the financial market, i.e., he divides his initial capital C_0 into A_0 shares of stock and B_0 shares of bonds. The bonds will yield a guaranteed interest rate r', whereas the stock price S_t, measured relative to time 0, is assumed to satisfy the growth equation (1.1). With a keen eye on the market the investor sells stocks to buy bonds and vice versa. Such dealings require no extra investments, and are called *self-financing*. Let A_t and B_t be the amounts of stocks and bonds held at time t. The total value C_t of stocks and bonds at time t is

$$C_t = A_t S_t + B_t e^{r't} . \tag{1.5}$$

The assumption that the tradings are self-financing can be expressed as:

$$dC_t = A_t dS_t + B_t d(e^{r't}). \tag{1.6}$$

An interesting question is now:

– What would our investor be prepared to pay at time 0 for a so-called *European call option*, i.e., the right to buy at some later time T a share of stock at a predetermined price K?

The rational answer, q say, was found by Black and Scholes (1973) through an analysis of the possible self-financing strategies leading from an initial investment q to the same payoff as the option would do. Their formula is now being used on stock markets all over the world.

The goal of this course is to first make sense of the above equations, and then to work with them.

1.2 What is noise?

In all the above examples the unexplained symbol N_t occurs, which is to be thought of as a "completely random" function of t, in other words, the continuous-time analogue of a sequence of independent identically distributed random variables.

In a first attempt to catch this concept, it is tempting to try and meet the following requirements:

1. N_t is independent of N_s for $t \neq s$.
2. The random variables N_t ($t \geq 0$) all have the same probability distribution μ.
3. $\mathbb{E}(N_t) = 0$.

However, these requirements do not produce what we want. We shall show that such a "continuous i.i.d. sequence" N_t is not measurable in t, unless it is identically zero.

Let μ denote the probability distribution of N_t, which by requirement 2 does not depend on t. If N_t is not a sure constant function of t, then there must be a value $a \in \mathbb{R}$ such that $p := \mathbb{P}[N_t \leq a]$ is neither 0 nor 1. Now consider the set E of time points where the noise is less than a. It can be shown that with probability 1 the set E is not Lebesgue measurable. Without giving a full proof we can understand this as follows.

Let λ denote the Lebesgue measure on $\mathbb{R}$. If E would be measurable, then by the independence requirement 1 we would expect its relative share in any interval (c,d) to be p, by a continuous analogue of the law of large numbers:

$$\lambda\left(E \cap (c,d)\right) = p\,(d-c)\ . \tag{1.7}$$

On the other hand, it is known from measure theory that every measurable set B is arbitrarily thick somewhere with respect to λ, i.e., for every $\alpha < 1$ an interval (c,d) can be found such that

$$\lambda\left(B \cap (c,d)\right) > \alpha(d-c)\ .$$

(cf. Halmos (1974) Theorem III.16.A: "Lebesgue's density theorem"). So by (1.7), E is not measurable. This is a bad property of N_t because in view of (1.1), (1.2), (1.3) and (1.4), we would like to be able to consider integrals of N_t.

For this reason, let us approach the problem from a different angle. Instead of N_t itself, let us consider directly the integral of N_t, and give it a name:

$$\text{“}B_t := \int_0^t N_s ds\text{”}.$$

The three requirements on the evasive object N_t then translate into three quite sensible requirements for B_t.

BM1. For any $0 = t_0 \le t_1 \le \ldots \le t_n$ the random variables $B_{t_{j+1}} - B_{t_j}$ $(j = 0, \ldots, n-1)$ are independent.

BM2. B_t has stationary increments, i.e., the joint probability distribution of

$$(B_{t_1+s} - B_{u_1+s}, B_{t_2+s} - B_{u_2+s}, \ldots, B_{t_n+s} - B_{u_n+s})$$

does not depend on $s \ge 0$, where $t_i > u_i \ge 0$ for $i = 1, 2, \ldots, n$ are arbitrary.

BM3. $\mathbb{E}\,(B_t) = 0$ for all $t \ge 0$.

We add a normalisation:

BM4. $\mathbb{E}(B_1^2) = 1$.

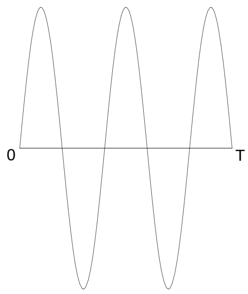
Still, these four requirements do not determine B_t. For example, the compensated Poisson jump process also satisfies them. Our fifth requirement fixes the process B_t uniquely (as will become clear later on):

BM5. $t \mapsto B_t$ is continuous a.s with probability 1.

The stochastic process B_t satisfying these requirements is called the *Wiener process* or *Brownian motion*. In the next chapter we shall give an explicit construction.

2 Brownian motion

In this section we shall construct Brownian motion on $[0, T]$ for some $T > 0$. We follow the original idea of Norbert Wiener in the 1930's. As we shall see, it clearly displays the noise N_t as a random *distribution* in the sense of Schwartz, rather than a random *function*.



Norbert Wiener **Fig. 0: the function e_n $(n = 5)$**

Consider the orthonormal basis $e_1, e_2, e_3, \dots$ of the Hilbert space $L^2[0, T]$ given by

$$e_n(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{n\pi t}{T}\right) \qquad (n \geq 1, \quad 0 \leq t \leq T).$$

For a function $f \in L^2[0, T]$ the coefficients on this basis (i.e. the Fourier coefficients) are the weights with which the frequencies $n = 1, 2, 3, \dots$ are represented in f. Now, 'white noise' should contain all frequencies with equal weights and in a random way. This leads to the following tentative definition of N_t:

$$\text{“ } N_t := \sum_{n=1}^{\infty} \omega_n e_n(t) \text{ ”} \quad (0 \leq t \leq T), \tag{2.1}$$

where $\omega_1, \omega_2, \omega_3, \dots$ is a sequence of independent standard Gaussian random variables. Now, the probability that the sequence $\omega_1, \omega_2, \omega_3, \dots$ should be square summable is 0. So N_t will almost surely not be an function in $L^2[0, T]$. But we shall be able to make sense of (2.1) in the sense of Schwartz distributions.

2.1 Notations

Let $\mathcal{S}_m$ $(m \in \mathbb{Z})$ denote the space of all sequences $\omega = (\omega_1, \omega_2, \dots) \in \mathbb{R}^{\mathbb{N}}$ for which

$$\|\omega\|_m^2 := \sum_{n=1}^{\infty} n^{2m} \omega_n^2 < \infty.$$

Note that $\mathcal{S}_0$ is the space $l^2(\mathbb{N})$ of square summable sequences. Let us denote by $\mathcal{S}$ the intersection

$$\mathcal{S} := \bigcap_{m \in \mathbb{Z}} \mathcal{S}_m,$$

and by $\mathcal{S}'$ the union

$$\mathcal{S}' := \bigcup_{m\in\mathbb{Z}} \mathcal{S}_m.$$

$\mathcal{S}$ consists of the rapidly decreasing sequences and $\mathcal{S}'$ of all the polynomially bounded ones. Under the pairing

$$\langle \omega, x\rangle := \sum_{n=1}^{\infty} \omega_n x_n,$$

the spaces $\mathcal{S}_m$ and $\mathcal{S}_{-m}$ are each others dual, and $\mathcal{S}'$ is the dual of $\mathcal{S}$. We have the following inclusions:

$$\mathcal{S}' \supset \cdots \supset \mathcal{S}_{-2} \supset \mathcal{S}_{-1} \supset \mathcal{S}_0 (= l^2) \supset \mathcal{S}_1 \supset \mathcal{S}_2 \supset \cdots \supset \mathcal{S}.$$

Next let $\mathbb{P}$ denote the infinite product measure on $\mathbb{R}^{\mathbb{N}}$ that makes $\omega = (\omega_1, \omega_2, \dots)$ into a sequence of independent standard Gaussian random variables. In a formula:

$$\mathbb{P}(d\omega) := \bigotimes_{n=1}^{\infty} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\omega_n^2} d\omega_n \right).$$

The following lemma says that $\mathbb{P}$-almost surely ω_n increases slower than n, making the square of their quotient summable.

Lemma 2.1 $\mathbb{P}(\mathcal{S}_{-1}) = \mathbb{P}\left\{ \omega \in \Omega : \sum_{n=1}^{\infty} \frac{\omega_n^2}{n^2} < \infty \right\} = 1.$

Proof. For $n \in \mathbb{N}$, let $M_n \colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ denote the random variable

$$M_n(\omega) := \sum_{j=1}^{n} \frac{\omega_j^2}{j^2}.$$

Since the ω_j's are standard Gaussian, we have for all $n \in \mathbb{N}$,

$$\mathbb{E}(M_n) = \sum_{j=1}^{n} \frac{1}{j^2} \le \frac{\pi^2}{6} =: c$$

and therefore, for all $k \in \mathbb{N}$ and $n \in \mathbb{N}$, by the Markov inequality,

$$\mathbb{P}[M_n > k] \le \frac{1}{k}\mathbb{E}(M_n) \le \frac{c}{k}.$$

Since $M_n(\omega)$ is increasing in n for all ω, it follows that

$$\mathbb{P}[M_n \le k \text{ for all } n] = \mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} [M_n \le k] \Big) = \lim_{n\to\infty} \mathbb{P}[M_n \le k] \ge 1 - \frac{c}{k},$$

and hence

$$\mathbb{P}(\mathcal{S}_{-1}) = \mathbb{P}\Big(\bigcup_{k\in\mathbb{N}}\bigcap_{n\in\mathbb{N}}[M_n \le k]\Big) = \lim_{k\to\infty}\Big(1-\frac{c}{k}\Big) = 1.$$

□

Since $\mathcal{S}_{-1} \subset \mathcal{S}'$, ω also lies in $\mathcal{S}'$ with probability 1. Now, $\mathcal{S}'$ consists of the sequences of Fourier coefficients of the tempered distributions, and the latter are a good class to work with. So we shall henceforth use $\mathcal{S}'$ as our sample space Ω. The connection with the space of distributions is as follows.

Let $\mathcal{D}[0,T]$ denote the space of infinitely differentiable functions on $[0,T]$ with compact support inside $(0,T)$. A function $f \in \mathcal{D}[0,T]$ determines a sequence of real numbers

$$\nu_n := \langle e_n, f\rangle := \int_0^T e_n(t)f(t)dt \quad (n \in \mathbb{N}).$$

For all even $m \in \mathbb{N}$, the property that f is m times continuously differentiable implies that $\nu \in \mathcal{S}_m$ by partial integration:

$$\begin{aligned}\int_0^T |f^{(m)}(t)|^2 dt &= \sum_{n=1}^{\infty}\langle e_n, f^{(m)}\rangle^2 = \sum_{n=1}^{\infty}\langle e_n^{(m)}, f\rangle^2 \\ &= \sum_{n=1}^{\infty}\Big(\frac{n\pi}{T}\Big)^{2m}\langle e_n, f\rangle^2 = \Big(\frac{\pi}{T}\Big)^{2m}\|\nu\|_m^2.\end{aligned}$$

Conversely, $\nu \in \mathcal{S}_m$ implies that $f^{(m)} \in L^2[0,t]$, so f is $m-1$ times continuously differentiable. Now, a sequence $\nu_1, \nu_2, \ldots$ is said to converge to an element ν of $\mathcal{S}$ (ν is itself a sequence of numbers!) if it converges in $\mathcal{S}_m$ for every m. On the other hand, a sequence $f_1, f_2, \ldots$ in $\mathcal{D}[0,T]$ is said to converge to an $f \in \mathcal{D}[0,T]$ if for all $m \in \mathbb{N}$ the sequence $f_1^{(m)}, f_2^{(m)}, \ldots$ converges uniformly to $f^{(m)}$. Therefore the topology on $\mathcal{S}$ is carried over to $\mathcal{D}[0,T]$.

2.2 Definition of Brownian motion

We now define *white noise* $\widetilde{N}$ on $[0,T]$ by

$$\widetilde{N}\colon \Omega \times \mathcal{D}[0,T]\colon (\omega, f) \mapsto \sum_{n=1}^{\infty}\omega_n\langle e_n, f\rangle.$$

For fixed $\omega \in \Omega$, the map $\widetilde{N}(\omega)\colon f \mapsto \widetilde{N}(\omega, f)$ is a *distribution* on $[0,T]$. Hence, with the measure $\mathbb{P}$ being defined on Ω, our map $\widetilde{N}$ becomes a *random distribution.* On the other hand, for fixed $f \in \mathcal{D}[0,T]$ the map

$$\widetilde{N}_f\colon \omega \mapsto \widetilde{N}(\omega, f)$$

is a Gaussian random variable with mean zero and variance $\|f\|^2$:

$$\begin{aligned}\mathbb{E}(\widetilde{N}_f) &= \mathbb{E}\Big(\sum_{i=1}^{\infty}\omega_i\langle e_i, f\rangle\Big) = 0 \\ \mathbb{E}(\widetilde{N}_f^2) &= \mathbb{E}\Big(\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\omega_i\omega_j\langle e_i, f\rangle\langle e_j, f\rangle\Big) = \sum_{i=1}^{\infty}\langle e_i, f\rangle^2 = \|f\|^2.\end{aligned}$$

It follows that the map $f \mapsto \widetilde{N}_f$ can be extended to a Hilbert space isometry $L^2[0,T] \to L^2(\Omega,\mathbb{P})$. Indeed, if f_i is a sequence of functions in $\mathcal{D}[0,T]$ such that $\|f_i - f\| \to 0$ $(i \to \infty)$ for some $f \in L^2[0,T]$, then the sequence $(\widetilde{N}_f)_{i=1}^\infty$ is a Cauchy-sequence in $L^2(\Omega,\mathbb{P})$:

$$\|\widetilde{N}_{f_i} - \widetilde{N}_{f_j}\|^2 = \mathbb{E}((\widetilde{N}_{f_i} - \widetilde{N}_{f_j})^2) = \mathbb{E}(\widetilde{N}^2_{f_i - f_j}) = \|f_i - f_j\|^2.$$

So it has a limit, N_f say. Moreover, the map $f \mapsto N_f$ is isometric:

$$\mathbb{E}(N_f^2) = \|f\|^2 \quad \forall f \in L^2[0,T]. \tag{2.2}$$

Hence again for all $f, g \in L^2[0,T]$:

$$\mathbb{E}(N_f) = 0 \;, \qquad \mathbb{E}(N_f N_g) = \langle f, g \rangle \;. \tag{2.3}$$

However, the extension from $\widetilde{N}$ on $\Omega \times \mathcal{D}[0,T]$ to N on $\Omega \times L^2[0,T]$ has a price: whereas $\widetilde{N}_f$ $(f \in \mathcal{D}[0,T])$ is a random element of $(\mathcal{D}[0,T])'$, N_f $(f \in L^2[0,T])$ is *not* a random element of $(L^2[0,T])' = L^2[0,T]$, the reason being that $N_f(\omega)$ is only defined for fixed $f \in L^2[0,T]$ and *almost* all $\omega \in \Omega$. The set of those ω for which $N_f(\omega)$ is defined for *all* $f \in L^2[0,T]$ has measure 0.

We employ the extension from $\widetilde{N}$ to N in order to define Brownian motion:

Definition. *Brownian motion over the time interval* $[0,T]$ is the family $(B_t)_{t\in[0,T]}$ of random variables $\Omega \to \mathbb{R}$ given by

$$B_t(\omega) := N_{1_{[0,t]}}(\omega).$$

We finally show that this definition meets all our requirements.

Proposition 2.2 *B_t satisfies the conditions* ***BM1–BM4*** *of Section 1. Moreover:*

BM5. *There exists a version of $(B_t)_{t\in[0,T]}$ such that $t \mapsto B_t$ is continuous.*

BM6. *With probability 1, $t \mapsto B_t$ has infinite variation over every time interval and is therefore nowhere differentiable.*

Proof. From the construction of N it is clear that for any $f_1, f_2, \ldots, f_k \in L^2[0,T]$ the random variables $N_{f_1}, N_{f_2}, \ldots, N_{f_k}$ are jointly Gaussian. A standard result in probability theory ensures that

i. jointly Gaussian random variables are independent as soon as they are uncorrelated,
ii. their joint probability distribution depends only on their expectations and their correlations.

So the independence of the increments of Brownian motion (**BM1**) follows from the orthogonality of the functions $1_{[t_j,t_{j+1})} = 1_{[0,t_{j+1})} - 1_{[0,t_j)}$ $(j = 0,\ldots,n-1)$, since N maps orthogonal functions f and g to uncorrelated random variables (by (2.3)). The stationarity of the increments (**BM2**) follows from the fact that the inner products of the functions $1_{[u_j+s,t_j+s)}$ $(j = 1,\ldots,n)$ do not depend on s. Properties **BM3** and **BM4** are obvious: $\mathbb{E}(B_t) = \mathbb{E}(N_{1_{[0,t]}}) = 0$, $\mathbb{E}(B_t^2) = \|1_{[0,t]}\|^2 = t$. The proof of properties **BM5** and **BM6** requires some preparation and is deferred to Sections 2.3–2.4. □

From (2.3) it follows that for all $s, t \geq 0$:

$$\mathbb{E}(B_s B_t) = \langle 1_{[0,s]}, 1_{[0,t]} \rangle = s \wedge t \;. \tag{2.4}$$

2.3 Continuity of paths and the problem of versions

The continuity of paths $t \mapsto B_t(\omega)$ with $\omega \in \Omega$ is a somewhat subtle matter.

First, let us make clear that the continuity of the curve $t \mapsto B_t$ in $L^2(\Omega, \mathbb{P})$ is not at all sufficient to ensure continuity of the paths $t \mapsto B_t(\omega)$.

Example 2.1.¡ Let $\Omega = [0, 1]$ with Lebesgue measure $\mathbb{P}$. Let X_t $(t \in [0, 1])$ denote the process

$$X_t(\omega) := 1_{[0,t]}(\omega) = \begin{cases} 0 & \text{if } t \leq \omega, \\ 1 & \text{if } t > \omega. \end{cases}$$

So $X_t(\omega)$, far from being continuous, moves by making a single jump at a random time. However, $t \mapsto X_t$ is a continuous curve in $L^2(\Omega, \mathbb{P})$, since

$$\|X_t - X_s\|^2 = \int_s^t du = t - s \quad (s \leq t) \, .$$

Second, let us observe that a description of a curve in $L^2(\Omega, \mathbb{P})$, such as $t \mapsto B_t$, can never be sufficient to ensure pathwise continuity. This is shown by the following example.

Example 2.2.¡ Take $(\Omega, \mathbb{P})$ as in the previous example. Now let $Y_t(\omega)$ be given by

$$Y_t(\omega) = \begin{cases} 1 & \text{if } \omega - t \text{ is irrational} \\ 0 & \text{if } \omega - t \text{ is rational.} \end{cases}$$

Let $Z_t(\omega) = 1$ for all $\omega \in \Omega$, $t \in [0, 1]$. Clearly, $t \mapsto Z_t(\omega)$ is constant, hence continuous for *all* $\omega \in \Omega$, whereas $t \mapsto Y_t(\omega)$ is discontinuous for all $\omega \in \Omega$. Nevertheless, for any fixed value of t, Y_t and Z_t are almost surely equal, so they correspond to the same curve in $L^2(\Omega, \mathbb{P})$.

These observations leads us to the following definition:

Definition. We say that a process $Y_t \in L^2(\Omega, \mathbb{P})$ $(t \in [0, T])$ is a *version* of a process $Z_t \in L^2(\Omega, \mathbb{P})$ $(t \in [0, T])$ if

$$\forall t \in [0, T] : \quad Y_t(\omega) = Z_t(\omega) \text{ for almost all } \omega \in \Omega.$$

This said, we proceed to prove that Brownian Motion has a version with continuous paths, as claimed in property **BM5** in Proposition 2.2.

Proof. For the fourth moment of a centered Gaussian random variable ξ we have $\mathbb{E}(\xi^4) = 3\,\mathrm{Var}(\xi)^2$. Hence for the process B_t we have

$$\mathbb{E}\big(|B_t - B_s|^4\big) = 3|t - s|^2.$$

Now fix $\alpha \in (0, 1/4)$ and put $\varepsilon := 1 - 4\alpha > 0$. Then, for all $s, t \in [0, T]$, we have by the Markov inequality,

$$\begin{aligned} \mathbb{P}\big[|B_t - B_s| \geq |t - s|^\alpha\big] &\leq |t - s|^{-4\alpha} \mathbb{E}\big(|B_t - B_s|^4\big) \\ &\leq 3|t - s|^{2-4\alpha} \\ &= 3|t - s|^{1+\varepsilon}, \end{aligned}$$

so that for all $n \in \mathbb{N}$ and $0 \leq k < 2^n$,

$$\mathbb{P}\big[|B_{(k+1)2^{-n}T} - B_{k2^{-n}T}| \geq 2^{-n\alpha}\big] \leq 3 \times 2^{-n(1+\varepsilon)},$$

and hence

$$\begin{aligned}
&\sum_{n=1}^{\infty}\sum_{k=0}^{2^n-1} \mathbb{P}\big[|B_{(k+1)2^{-n}T} - B_{k2^{-n}T}| \geq 2^{-n\alpha}\big] \\
&\leq \sum_{n=1}^{\infty}\sum_{k=0}^{2^n-1} 3 \times 2^{-n}2^{-n\varepsilon} = 3\sum_{n=1}^{\infty} 2^{-n\varepsilon} = \frac{3}{2^{\varepsilon}-1} < \infty.
\end{aligned}$$

By the first Borel-Cantelli lemma it follows that if A_n denotes the event

$$A_n := \big[\exists 0 \leq k < 2^n \colon |B_{(k+1)2^{-n}T} - B_{k2^{-n}T}| \geq 2^{-n\alpha}\big],$$

then with probability 1 only finitely many A_n's happen. So for almost all $\omega \in \Omega$ there exists $M(\omega) \in \mathbb{N}$ such that A_n does not occur for $n \geq M(\omega)$, i.e., for all $n \geq M(\omega)$ we have

$$\forall 0 \leq k < 2^n \colon |B_{(k+1)2^{-n}T}(\omega) - B_{k2^{-n}T}(\omega)| < 2^{-n\alpha}. \tag{2.5}$$

Now fix an ω for which the latter holds, and let s and t be two points in the set Q of dyadic rationals of T,

$$Q := \left\{ k2^{-n}T \;\middle|\; n, k \in \mathbb{N}, 0 \leq k \leq 2^n \right\},$$

such that $|s-t| \leq 2^{-M(\omega)}T$. Choose $n \geq M(\omega)$ such that $2^{-(n+1)}T \leq |s-t| < 2^{-n}T$. Let $k \in \mathbb{N}$ be such that $k2^{-n}T$ has distance less than $2^{-n}T$ to both s and t. We may then write the dyadic representation

$$\begin{aligned}
s &= k2^{-n}T + \sum_{i=1}^{l} \sigma_i 2^{-(n+i)}T \\
t &= k2^{-n}T + \sum_{j=1}^{m} \tau_i 2^{-(n+j)}T,
\end{aligned}$$

where $l, m \in \mathbb{N}$ and $\sigma_i, \tau_j \in \{-1, 0, 1\}$ for $i = 1, \ldots, l$ and $j = 1, \ldots, m$. By (2.5) we may now conclude that

$$\begin{aligned}
|B_t(\omega) - B_s(\omega)| &< \sum_{i=1}^{l} |\sigma_i 2^{-(n+i)\alpha}| + \sum_{j=1}^{m} |\tau_j 2^{-(n+j)\alpha}| \\
&\leq \frac{2}{1-2^{-\alpha}} 2^{-(n+1)\alpha} \\
&\leq \frac{2}{T^{\alpha}(1-2^{-\alpha})}|s-t|^{\alpha}.
\end{aligned}$$

It follows that the restriction of the function $t \mapsto B_t(\omega)$ to Q may be extended to a continuous function $t \mapsto \widetilde{B}_t(\omega)$ on all of $[0, T]$.

It remains to show that $\widetilde{B}_t$ is a version of B_t. Choose $t \in [0,T]$ and let $t_1, t_2, \ldots$ be a sequence in Q tending to t. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a bounded, continuous and strictly increasing function. Then

$$\int_\Omega |\varphi(B_t(\omega)) - \varphi(\widetilde{B}_t(\omega))|^2 \mathbb{P}(d\omega) = \|\varphi(B_t) - \varphi(\widetilde{B}_t)\|^2 = \lim_{i\to\infty} \|\varphi(B_{t_i}) - \varphi(\widetilde{B}_{t_i})\|^2 = 0,$$

since $t \mapsto B_t$ is continuous in $L^2(\Omega, \mathbb{P})$ and $t \mapsto \widetilde{B}_t$ is pointwise continuous, and therefore also continuous in $L^2(\Omega, \mathbb{P})$ by bounded convergence. □

Henceforth, when speaking of B_t we shall always mean its continuous version and drop the tilde from the notation.

2.4 Roughness of paths

Although the path of Brownian motion is continuous, it is extremely rough. This is expressed by the following basic lemma, which we shall need to prove property **BM6** in Proposition 2.2.

Lemma 2.3 *Let* $0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = T$ *be a family of partitions of the interval* $[0,T]$ *that gets arbitrarily fine for* $n \to \infty$*, in the sense that*

$$\lim_{n\to\infty} \max_{0\le j<n} |t_{j+1}^{(n)} - t_j^{(n)}| = 0.$$

Then

$$L^2 - \lim_{n\to\infty} \sum_{j=0}^{n-1} (B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}})^2 = T.$$

Proof. Abbreviate $\Delta B_j = B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}}$ and $\Delta t_j = t_{j+1}^{(n)} - t_j^{(n)}$. Let $\delta_n = \max_j \Delta t_j$. Then

$$\begin{aligned}
\Big\|\sum_j (\Delta B_j)^2 - T\Big\|^2 &= \mathbb{E}\Big[\Big(\sum_j (\Delta B_j)^2 - T\Big)^2\Big] \\
&= \mathbb{E}\Big(\sum_{i,j} (\Delta B_i)^2 (\Delta B_j)^2\Big) - 2T\mathbb{E}\Big(\sum_j (\Delta B_j)^2\Big) + T^2 \\
&= \sum_j \mathbb{E}((\Delta B_j)^4) + \sum_{i\ne j} \mathbb{E}((\Delta B_i)^2)\mathbb{E}((\Delta B_j)^2) - 2T\Big(\sum_j \Delta t_j\Big) + T^2 \\
&= \sum_j 3(\Delta t_j)^2 + \sum_{i\ne j} (\Delta t_i)(\Delta t_j) - T^2 \\
&= 2\sum_j (\Delta t_j)^2 \\
&\le 2\delta_n \sum_j \Delta t_j \\
&= 2\delta_n T,
\end{aligned}$$

where again we have used the fact that the fourth moment of a centered Gaussian random variable ξ is given by $\mathbb{E}\left(\xi^4\right) = 3\,\mathrm{Var}(\xi)^2$. As $\lim_{n\to\infty}\delta_n = 0$, the statement follows. □

We may write the message of Lemma 2.3 symbolically as

$$\text{“}(dB_t)^2 = dt\text{”}, \tag{2.6}$$

saying that “Brownian motion has quadratic variation growing linearly with time”. This expression will acquire a precise meaning during the sequel of this course. For the moment, let us just say that B_t has large fluctuations on a small scale, namely

$$dB_t \text{ is of order } \sqrt{dt} \ \left(\gg dt\right).$$

To prove **BM6** of Proposition 2.2 we need one more preparatory lemma, which applies to a general sequence of random variables.

Lemma 2.4 *Let $X_1, X_2, X_3, \dots$ be a sequence of real-valued random variables such that for some $p > 0$,*

$$\lim_{n\to\infty} \mathbb{E}\left(|X_n|^p\right) = 0 \quad \textit{for some } p > 0.$$

Then there exists a subsequence $(X_{n_k})_{k=1}^{\infty}$ tending to 0 almost surely.

Proof. Choose a subsequence such that $\sum_{k=1}^{\infty}\mathbb{E}(|X_{n_k}|^p) < \infty$. By Chebyshev's inequality we have, for all $m \in \mathbb{N}$,

$$\mathbb{P}\left[|X_{n_k}| \geq \frac{1}{m}\right] \leq m^p\mathbb{E}\left(|X_{n_k}|^p\right),$$

and therefore, for all $m \in \mathbb{N}$,

$$\sum_k \mathbb{P}\left[|X_{n_k}| \geq \frac{1}{m}\right] < \infty.$$

The first Borel-Cantelli lemma now implies that for any m,

$$\mathbb{P}\left[|X_{n_k}| \geq \frac{1}{m} \text{ for finitely many } k\right] = 1.$$

By σ-additivity the intersection over m of the event between brackets also has probability 1:

$$\mathbb{P}\left[\forall m \in \mathbb{N}\,\exists K \in \mathbb{N}\,\forall k \geq K\colon\ |X_{n_k}| < \frac{1}{m}\right] = 1,$$

i.e., $X_{n_k} \to 0$ almost surely as $k \to \infty$. □

We are finally ready to finish our proof of Proposition 2.2:

Proof. By Lemmas 2.3 and 2.4 (with $p = 2$) there exists an increasing sequence (n_k) such that, for almost all $\omega \in \Omega$,

$$\lim_{k\to\infty} \sum_{j=0}^{n-1} \left(B_{t_{j+1}^{(n_k)}}(\omega) - B_{t_j^{(n_k)}}(\omega) \right)^2 = T.$$

Fix an $\omega \in \Omega$ for which this holds. Let

$$\varepsilon_{n_k} := \max_j |\Delta B_j| := \max_{0\le j<n_k} \left| B_{t_{j+1}^{(n_k)}}(\omega) - B_{t_j^{(n_k)}}(\omega) \right| .$$

Then $\lim_{k\to\infty} \varepsilon_{n_k} = 0$ by the uniform continuity of $t \mapsto B_t$ (which is a continuous function on the compact interval $[0, T]$). It follows that,

$$\sum_{j=0}^{n_k-1} |\Delta B_j| \ge \sum_{j=0}^{n_k-1} \frac{1}{\varepsilon_{n_k}} |\Delta B_j|^2 \sim \frac{1}{\varepsilon_{n_k}} T \longrightarrow \infty \quad \text{as} \quad k \to \infty$$

□

This concludes the construction of Brownian motion. In the next sections we shall see that Brownian motion is the building block of stochastic analysis.

The construction of Brownian motion induces a probability measure on the set of all continuous functions $[0, T] \to \mathbb{R}$, which is called the *Wiener measure.*

It can be proved that *all* random variables in $L^2(\Omega, \mathbb{P})$ can be represented in a natural way as integrals of products of increments of Brownian motion. This is Wiener's *chaos expansion*, (Wiener (1938).)

3 Stochastic integration: the Gaussian case

This section serves as a motivation for the Itô-calculus presented in Sections 4 and 5.

3.1 The stochastic integral of a non-random function

If $f_n \in L^2[0,T]$ is a step function, i.e.,

$$f_n(t) = \sum_{j=0}^{n-1} c_j^{(n)} 1_{[t_j,t_{j+1})}(t) \qquad (0 = t_0 < t_1 < \cdots < t_n = T),$$

then from the definition $B_t := N_{1_{[0,T]}}$ it follows that

$$N_{f_n} = \sum_{j=0}^{n-1} c_j^{(n)} (B_{t_{j+1}} - B_{t_j}),$$

which is precisely what we would mean by the integral $\int_0^T f_n(t)dB_t$. Moreover, if $f_n \to f$ in $L^2[0,T]$, then $N_{f_n} \to N_f$ in $L^2(\Omega,\mathbb{P})$ by the isometric property of N described in Section 2.2. So $f \mapsto N_f$ maps $L^2[0,T]$ into the Gaussian random variables on $(\Omega,\mathbb{P})$. It is therfore natural to define for all $f \in L^2[0,T]$:

$$\int_0^T f(t)dB_t(\omega) := N_f(\omega).$$

3.2 The Ornstein-Uhlenbeck process

In 1908 Langevin proposed an equation in order to describe the motion of a particle suspended in a liquid. He wanted to give a treatment which was more refined and more in harmony with Newtonian mechanics than the above Brownian motion model. Let V_t denote the velocity of the particle at time t. According to Newton's second law the derivative $\frac{d}{dt}V_t$ should be equal to the sum of the forces acting on the particle (see example 2 in Section 1: we take the mass of the particle equal to 1). Langevin wrote

$$\frac{d}{dt}V_t = -\eta V_t + \sigma N_t,$$

where the first term on the r.h.s. models the friction due to the viscosity η of the liquid and the second term describes the sum total of the erratic collisions of liquid molecules against the particle, σ being the strength of the noise.

In the spirit of the preceding discussion we rewrite the Langevin equation as

$$dV_t = -\eta V_t dt + \sigma dB_t, \tag{3.1}$$

which again is shorthand for the integral equation

$$V_t - V_s = -\eta \int_s^t V_u du + \sigma(B_t - B_s) \tag{3.2}$$

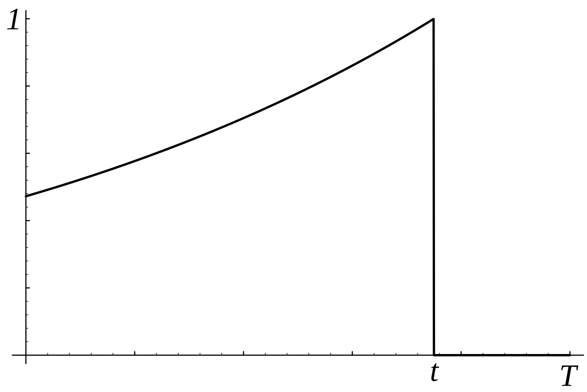


Figure 1: The function f_t.

in $L^2(\Omega, \mathbb{P})$. This equation was solved by Ornstein and Uhlenbeck (1930) in the following way. Rewrite (3.1) as

$$d(e^{\eta t} V_t) = \sigma e^{\eta t} dB_t.$$

Integrating we obtain, for $0 \le t \le T$,

$$e^{\eta t} V_t - V_0 = \sigma \int_0^t e^{\eta s} dB_s,$$

so we find

$$\begin{aligned} V_t &= e^{-\eta t} V_0 + \sigma \int_0^t e^{-\eta(t-s)} dB_s \\ &= e^{-\eta t} V_0 + \sigma N_{f_t}, \end{aligned} \tag{3.3}$$

where $f_t \colon s \mapsto 1_{[0,t]}(s) \exp[-\eta(t-s)]$ is shown in Figure 1.

Assuming $V_0 = 0$, we obtain the following first and second moments for the Gaussian random variable V_t (recall (2.2)):

$$\mathbb{E}(V_t) = 0, \qquad \mathbb{E}(V_t^2) = \sigma^2 \|f_t\|^2 = \sigma^2 \int_0^t e^{2\eta(s-t)} ds = \frac{\sigma^2}{2\eta}(1 - e^{-2\eta t}).$$

Ignoring for the moment that time stops at T (T is arbitrary and could be taken very large), we take the limit $t \to \infty$:

$$\lim_{t \to \infty} \mathbb{E}(V_t^2) = \frac{\sigma^2}{2\eta}. \tag{3.4}$$

This can be read as saying that the particle will eventually attain a certain mean energy.

Let us now show that the above informal calculation is indeed correct.

Proposition 3.1 *For all $V_0 \in \mathbb{R}$ the process $(V_t)_{t \in [0,T]}$ defined by (3.3) satisfies the stochastic integral equation (3.2).*

Proof. After substitution of (3.3) into (3.2), the latter is found to consist of a deterministic part

$$e^{-\eta t}V_0 - e^{-\eta s}V_0 = -\eta \int_s^t e^{-\eta u}V_0 du,$$

which is obviously valid, and a stochastic part

$$\sigma N_{f_t} - \sigma N_{f_s} = -\eta\sigma \int_s^t N_{f_u} du + \sigma(B_t - B_s),$$

which can be obtained by applying the map N to both sides of the following identity in $L^2[0,T]$:

$$f_t - f_s = -\eta \int_s^t f_u du + 1_{[s,t]} \qquad (0 \leq s \leq t \leq T).$$

The proof of this identity is left as an exercise. □

3.3 Some hocus pocus

Let us again consider the Ornstein-Uhlenbeck process $(V_t)_{t\in[0,T]}$ and perform some suggestive formal manipulations.

It is tempting to write

$$d(V_t^2) = 2V_t dV_t = 2V_t(-\eta V_t dt + \sigma dB_t),$$

which, under the formal assumption that V_t and dB_t are independent quantities, leads to

$$\begin{aligned} d\mathbb{E}(V_t^2) &= -2\eta\mathbb{E}(V_t^2)dt + 2\sigma\mathbb{E}(V_t)\mathbb{E}(dB_t) \\ &= -2\eta\mathbb{E}(V_t^2)dt. \end{aligned}$$

However, this result is false since it would imply that $\mathbb{E}(V_t^2)$ tends to 0 as $t \to \infty$, contradicting (3.4).

What went wrong? How should we change the rules of formal manipulation of differentials? Should we reject the independence assumption $V_t \perp\!\!\!\perp dB_t$? This is indeed a possibility, and leads to the so-called "stochastic calculus of Stratonovic". However, we shall not reject this independence assumption, but instead follow Itô and reconsider the first step $d(V_t^2) = 2V_t dV_t$.

Let us take seriously the formula $(dB_t)^2 = dt$ in (2.6), implying that $(dV_t)^2 = \sigma^2 dt$, and let us expand $d(V_t^2)$ to second order in dV_t:

$$\begin{aligned} d(V_t^2) &= 2V_t dV_t + (dV_t)^2 \\ &= (V_t + dV_t)^2 - V_t^2 \\ &= 2V_t(-\eta V_t dt + \sigma dB_t) + \sigma^2 dt. \end{aligned}$$

This gives, upon calculation of expectations,

$$d\mathbb{E}(V_t^2) = -2\eta\mathbb{E}(V_t^2)dt + \sigma^2 dt,$$

leading indeed to the correct equilibrium value (3.4). Note the occurrence of the extra term $\sigma^2 dt$ compared to the previous calculation. This term is the essential ingredient to make ends meet.

The above argument is heuristic at this stage, because both B_t and V_t are functions of infinite variance, and hence nowhere differentiable. Still, we shall see that it makes perfect sense in the right interpretation. This will be clarified in Sections 4 and 5.

4 The Itô-integral

Kiyosi Itô

In this section we shall extend the concept of stochastic integration by allowing the function $f(t,\omega)$ in the integral $\int_0^T f(t,\omega)dB_t(\omega)$ to become stochastic as well. However, we shall see that $f(t,\omega)$ cannot be completely arbitrary, but has in some way to be fitting to $\omega \mapsto B_t(\omega)$. The construction was pioneered by Kiyosi Itô in the 1940's and leads to what is nowadays called *Itô's stochastic calculus.*

4.1 Step functions

We define the stochastic integral of a step function of the form

$$\phi(t,\omega) = \sum_{j=0}^{n-1} c_j(\omega) 1_{[t_j,t_{j+1})}(t)$$

by

$$\int_0^T \phi(t,\omega)dB_t(\omega) := \sum_{j=0}^{n-1} c_j(\omega)\big(B_{t_{j+1}}(\omega) - B_{t_j}(\omega)\big). \tag{4.1}$$

The next thing to do would be to approximate f by step functions f_n and define $\int f dB_t$ to be the limit of $\int f_n dB_t$. But here we meet a difficulty!

Example 4.1.¡ Put $f(t,\omega) := B_t(\omega)$. Two reasonable approximations of f are ϕ_n and ψ_n given by

$$\phi_n(t,\omega) := \sum_{j=0}^{n-1} B_{t_j}(\omega) 1_{[t_j,t_{j+1})}(t),$$

$$\psi_n(t,\omega) := \sum_{j=0}^{n-1} B_{t_{j+1}}(\omega) 1_{[t_j,t_{j+1})}(t),$$

where $t_0, t_1, \ldots, t_n$ are defined as in Lemma 2.3 in Section 2.4 (and are not ω-dependent). However, from our definition (4.1) we find that

$$\int_0^T \psi_n dB_t - \int_0^T \phi_n dB_t = \sum_{j=0}^{n-1} (\Delta B_j)^2,$$

which, according to Lemma 2.3, does not tend to 0 as $n \to \infty$ but to the constant T. In other words, the variation of the path $t \mapsto B_t$ is too large for $\int B_t dB_t$ to be defined in a casual way.

We now introduce a requirement on our function f for the approximation by step functions f_n to work nicely.

Definition. Let $\mathcal{F}_t$ denote the σ-field generated by $\{ B_s \mid 0 \le s \le t \}$. Let $\mathcal{F} := \mathcal{F}_T$. A stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of Brownian motion is a measurable map $[0,T] \times \Omega \to \mathbb{R}$. The process will be called *adapted* to the family of σ-fields $(\mathcal{F}_t)_{t \in [0,T]}$ if $\omega \mapsto f(t,\omega)$ is $\mathcal{F}_t$-measurable for all $t \in [0,T]$.

The space $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$ of square-integrable $\mathcal{F}_t$-measurable functions may be thought of as those functions of $\omega \in \Omega$ that are fully determined by the initial segment $[0,t] \to \mathbb{R}$: $s \mapsto B_s(\omega)$ of the Brownian motion. In other words, $g \in L^2(\Omega, \mathcal{F}_t, \mathbb{P})$ when $g(\omega) = g(\omega')$ as soon as $B_s(\omega) = B_s(\omega')$ for all $s \in [0,t]$.

Let $\mathcal{L}^2(B, [0,T])$ denote the space of all adapted stochastic processes $f\colon [0,T] \times \Omega \to \mathbb{R}$ that are square integrable:

$$\|f\|^2_{L^2(\Omega,\mathbb{P})} = \int_0^T dt \int_\Omega f(t,\omega)^2 \mathbb{P}(d\omega) < \infty.$$

The natural inner product that makes $\mathcal{L}^2(B, [0,T])$ into a real Hilbert space is

$$\begin{aligned}\langle f, g \rangle &:= \int_0^T dt \int_\Omega f(t,\omega) g(t,\omega) \mathbb{P}(d\omega) \\ &= \mathbb{E}\Big(\int_0^T f(t,\cdot) g(t,\cdot) dt \Big).\end{aligned}$$

We note that the step functions ϕ_n in the last example are adapted, since $\phi_n(t,\omega) = B_{t_j}$ for $t \in [t_j, t_{j+1})$, so that $\phi_n(t,\omega)$ only depends on past values of B. On the other hand, ψ_n is not adapted, since at time $t \in [t_j, t_{j+1})$ it already anticipates the Brownian motion at time t_{j+1}: $\psi_n(t,\omega) = B_{t_{j+1}}(\omega)$.

The next theorem is a crucial property of stochastic integrals of adapted step functions.

Proposition 4.1 (The Itô-isometry) *Let ϕ be a step function in $\mathcal{L}^2(B, [0,T])$, and let*

$$\mathcal{I}_0(\phi)(\omega) := \int_0^T \phi(t,\omega) dB_t(\omega)$$

be its stochastic integral according to (4.1). Then $\mathcal{I}_0$ is an isometry:

$$\|\mathcal{I}_0(\phi)\|_{L^2(\Omega,\mathbb{P})} = \|\phi\|_{\mathcal{L}^2(B)}, \tag{4.2}$$

i.e.,

$$\int_\Omega \mathbb{P}(d\omega)\left(\int_0^T \phi(t,\omega)dB_t(\omega)\right)^2 = \int_0^T dt \int_\Omega \mathbb{P}(d\omega)\,\phi^2(t,\omega).$$

Proof. By adaptedness, c_i in (4.1) is *independent* of $\Delta B_j := B_{t_{j+1}} - B_{t_j}$ for $0 \le i \le j$. Therefore

$$\begin{aligned}
\|\mathcal{I}_0(\phi)\|^2_{L^2(\Omega,\mathbb{P})} &= \mathbb{E}\Big(\Big(\sum_{j=0}^{n-1} c_j \Delta B_j\Big)^2\Big)\\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \mathbb{E}\big(c_i c_j \Delta B_i \Delta B_j\big)\\
&= \sum_{j=0}^{n-1} \mathbb{E}\big(c_j^2 (\Delta B_j)^2\big) + 2\sum_{i<j} \mathbb{E}\big(c_i c_j \Delta B_i\big)\mathbb{E}\big(\Delta B_j\big)\\
&= \sum_{j=0}^{n-1} \mathbb{E}(c_j^2)\mathbb{E}\big((\Delta B_j)^2\big)\\
&= \sum_{j=0}^{n-1} \mathbb{E}(c_j^2)\Delta t_j,
\end{aligned}$$

where we use that $\mathbb{E}(\Delta B_j) = 0$, $\mathbb{E}((\Delta B_j)^2) = \Delta t_j$ (recall **BM3-BM4** in Section 1). On the other hand,

$$\begin{aligned}
\|\phi\|^2_{\mathcal{L}(B,[0,T])} &= \int_0^T \mathbb{E}\Big(\Big[\sum_{j=0}^{n-1} c_j 1_{[t_j,t_{j+1})}(t)\Big]^2\Big)dt\\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\Big(\int_0^T 1_{[t_i,t_{i+1})}(t) 1_{[t_j,t_{j+1})}(t)dt\Big)\mathbb{E}(c_i c_j)\\
&= \sum_{j=0}^{n-1} \Delta t_j \mathbb{E}(c_j^2).
\end{aligned}$$

The two expressions are the same. □

4.2 Arbitrary functions

To go from step functions to arbitrary functions we need the following.

Lemma 4.2 *Every function $f \in \mathcal{L}^2(B,[0,T])$ can be approximated arbitrarily well by step functions in $\mathcal{L}^2(B,[0,T])$.*

Proof of Lemma 4.2. We divide the proof into three steps of successive approximation.

Step 1. Every bounded pathwise continuous $g \in \mathcal{L}^2(B,[0,T])$ can be approximated by a sequence of step functions.

Proof. Partition the interval $[0,T]$ into n pieces by times $(t_j)_{j=1}^n$ in the customary way. Define

$$\phi_n(t,\omega) := \sum_{j=0}^{n-1} g(t_j,\omega)1_{[t_j,t_{j+1})}(t).$$

Then, since $t \mapsto g(t,\omega)$ is continuous and $\max_j |\Delta t_j| \to 0$ for all $\omega \in \Omega$, we have

$$\lim_{n\to\infty} \int_0^T \big(g(t,\omega) - \phi_n(t,\omega)\big)^2 dt = 0.$$

Hence, by bounded convergence,

$$\lim_{n\to\infty} \mathbb{E}\Big(\int_0^T \big(g(t,\omega) - \phi_n(t,\omega)\big)^2 dt\Big) = 0.$$

□

Step 2. Every bounded $h \in \mathcal{L}^2(B)$ can be approximated by a sequence of bounded continuous functions in $\mathcal{L}^2(B)$.

Proof. Suppose $|h| \le M$. For $n \in \mathbb{N}$, let ψ_n be a non-negative continuous function of the form given in Figure 2, with the properties $\psi_n(x) = 0$ for $x \notin [0,1/n]$ and $\int_{-\infty}^{\infty} \psi_n(x)dx = 1$. Define

$$g_n(t,\omega) := \int_0^t \psi_n(t-s)h(s,\omega)ds.$$

(Think of ψ_n as a "mollifier" of h.) Then $t \mapsto g_n(t,\omega)$ is continuous for all ω, and $|g_n| \le M$. Moreover, for all ω,

$$\lim_{n\to\infty} \int_0^T \big(g_n(s,\omega) - h(s,\omega)\big)^2 ds = 0,$$

since ψ_n constitutes an approximate identity. Again, by bounded convergence,

$$\lim_{n\to\infty} \mathbb{E}\Big(\int_0^T \big(g_n(s,\omega) - h(s,\omega)\big)^2 ds\Big) = 0.$$

□

Step 3. Every $f \in \mathcal{L}^2(B,[0,T])$ can be approximated by bounded functions in $\mathcal{L}^2(B,[0,T])$. (This is in fact a general property L^2-spaces.)

Proof. Let $f \in \mathcal{L}^2(B)$ and put $h_n(t,\omega) := (-n) \vee (n \wedge f(t,\omega))$. Then

$$\|f - h_n\|^2_{\mathcal{L}^2(B)} \le \int_0^T dt \int_\Omega \mathbb{P}(d\omega)\; 1_{[n,\infty)}(|f(t,\omega)|)f(t,\omega)^2,$$

which tends to 0 as $n \to \infty$ by dominated convergence. □

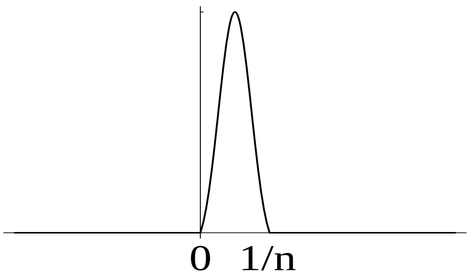


Figure 2: The function ψ_n.

On the basis of Proposition 4.1 and Lemma 4.2 we can now define the Itô-integral of a function $g \in \mathcal{L}^2(B, [0,T])$ as follows. Approximate g by step functions $\phi_n \in \mathcal{L}^2(B, [0,T])$, i.e., $\phi_n \to g$ in $\mathcal{L}^2(B, [0,T])$. Apply $\mathcal{I}_0$ to each of the ϕ_n. Since $\mathcal{I}_0$ is an isometry, the sequence $\mathcal{I}_0\phi_n$ has a limit in $L^2(\Omega, \mathbb{P})$. This is what we *define* to be the Itô-integral $\mathcal{I}g$ of g:

$$\int_0^T g(t,\omega)dB_t(\omega) := (\mathcal{I}g)(\omega) = L^2 - \lim_{n\to\infty} (\mathcal{I}_0\phi_n)(\omega).$$

Here is an example of a stochastic integral.

Example 4.2.¡ The following identity holds:

$$\int_0^T B_t dB_t = \frac{1}{2}B_T^2 - \frac{1}{2}T.$$

Proof. We choose an adapted approximation of $B_t(\omega)$, namely $\phi_n(t,\omega)$ of the example in Section 4.1. By definition,

$$\int_0^T \phi_n(t)dB_t = \sum_{j=0}^{n-1} B_j \Delta B_j,$$

where we use the shorthand notation $B_j := B_{t_j}$ and $\Delta B_j := B_{t_{j+1}} - B_{t_j}$. Note that $B_i = \sum_{0\le j<i} \Delta B_j$. We therefore have

$$\begin{aligned} B_T^2 &= \Big(\sum_j \Delta B_j\Big)^2 \\ &= \sum_i (\Delta B_i)^2 + 2\sum_{i<j} (\Delta B_i)(\Delta B_j) \\ &= \sum_i (\Delta B_i)^2 + 2\sum_j B_j(\Delta B_j) \\ &= \sum_i (\Delta B_i)^2 + 2\int_0^T \phi_n(t)dB_t. \end{aligned}$$

From Lemma 2.3 in Section 2.4 it now follows that

$$\mathcal{I}B = \int_0^T B_t dB_t = \lim_{n\to\infty} \int_0^T \phi_n(t)dB_t = \frac{1}{2}(B_T^2 - T).$$

□

Note that the integral in the above example is actually *different* from what it would be for a smooth function f with $f(0) = 0$, namely: $\int_0^T f(t)df(t) = \frac{1}{2}f(T)^2$. What the example shows is that "stochastic integration is ordinary integration except that the diagonal terms must be left out". This will be made precise in Section 5, where we shall encounter a faster way to calculate stochastic integrals.

4.3 Martingales

In Section 4.4 we shall prove that the Itô-integral w.r.t. Brownian motion of an adapted square-integrable stochastic process always has a continuous version. For this we shall need the interlude on martingales described in this section. For a general introduction on martingales we refer to the book by D. Williams (1991).

Definition. By the conditional expectation at time $t \in [0,T]$ of a random variable $X \in L^2(\Omega, \mathbb{P})$ we shall mean its orthogonal projection onto $L^2(\Omega, \mathcal{F}_t, \mathbb{P})$, the space of random variables that are determined by the Brownian motion up to time t. We denote this projection by $\mathbb{E}(X \mid \mathcal{F}_t)$, or briefly $\mathbb{E}_t(X)$.

In words, $\mathbb{E}_t(X)(\omega)$ is the best estimate (in the sense of least mean square error) that can be made of $X(\omega)$ on the basis of the knowledge of $B_s(\omega)$ for $0 \leq s \leq t$.

Definition. An adapted process $M \in \mathcal{L}^2(B,[0,T])$ is called a *martingale* (w.r.t. Brownian motion) if

$$\mathbb{E}_s(M_t) = M_s \qquad \text{for } 0 \leq s \leq t \leq T.$$

A martingale is a "fair game": the expected value at any time in the future is equal to the current value. Note that Brownian motion itself is a martingale, since for $0 \leq s \leq t \leq T$,

$$\mathbb{E}_s(B_t) = \mathbb{E}_s(B_s + (B_t - B_s)) = B_s + \mathbb{E}_s(B_t - B_s) = B_s,$$

because $B_t - B_s$ is independent of and hence orthogonal to any function in $L^2(\Omega, \mathcal{F}_s, \mathbb{P})$, in particular, to Brownian motion.

Theorem 4.3 *The stochastic integral of an adapted step function is a martingale with continuous paths.*

Proof. This follows directly from the fact that Brownian motion has continuous paths and satisfies the martingale property (use the definition of the stochastic integral of a step function given in (4.1) in Section 4.1). □

The following powerful tool will help us prove that the Itô-integral of *any* process in $\mathcal{L}^2(B,[0,T])$ possesses a continuous version.

Theorem 4.4 (The Doob martingale inequality) *If M_t is a martingale with continuous paths, then for all $p \geq 1$ and $\lambda > 0$,*

$$\mathbb{P}[\sup_{0 \leq t \leq T} |M_t| \geq \lambda] \leq \frac{1}{\lambda^p}\mathbb{E}(|M_T|^p).$$

Proof. We may assume that $\mathbb{E}(|M_T|^p) < \infty$, otherwise the statement is trivially true. Let $Z_t := |M_t|^p$. Then, since $x \mapsto |x|^p$ is a convex function, Jensen's inequality gives that Z_t is sub-martingale: for all $0 \le s \le t \le T$,

$$\mathbb{E}_s(Z_t) = \mathbb{E}_s(|M_t|^p) \ge |\mathbb{E}_s(M_t)|^p = |M_s|^p = Z_s.$$

It follows in particular that $\mathbb{E}(|M_s|^p) < \infty$ for all $s \in [0,T]$. Let us discretise time and first prove a discrete version of the Doob inequality. To that end we fix $n \in \mathbb{N}$ and put $t_k = kT/n$, $k = 0, 1, \dots, n$. Let $K(\omega)$ denote the smallest value of k for which $Z_{t_k} \ge \lambda^p$, if this occurs at all. Otherwise, put $K(\omega) = \infty$. Then we may write, since $[K = k] \in \mathcal{F}_{t_k}$,

$$\begin{aligned}
\mathbb{P}[\max_{0\le k\le n} |M_{t_k}| \ge \lambda] &= \sum_{k=0}^{n} \mathbb{P}[K = k] \\
&\le \sum_{k=0}^{n} \frac{1}{\lambda^p} \mathbb{E}(1_{[K=k]} Z_{t_k}) \\
&\le \sum_{k=0}^{n} \frac{1}{\lambda^p} \mathbb{E}\big(1_{[K=k]} \mathbb{E}(Z_T | \mathcal{F}_{t_k})\big) \\
&= \sum_{k=0}^{n} \frac{1}{\lambda^p} \mathbb{E}(1_{[K=k]} Z_T) \\
&\le \frac{1}{\lambda^p} \mathbb{E}(Z_T) \\
&= \frac{1}{\lambda^p} \mathbb{E}(|M_T|^p).
\end{aligned}$$

Here, the second inequality uses the sub-martingale property at time t_k. To get the same for continuous time, let A_n denote the event $[\max_{0\le k\le n} |M_{t_k}| \ge \lambda]$. Then we have $A_1 \subset A_2 \subset A_4 \subset A_8 \subset \cdots$, and so, $t \mapsto M_t$ being continuous,

$$\mathbb{P}[\sup_{0\le t\le T} |M_t| \ge \lambda] = \mathbb{P}\big(\bigcup_{n=0}^{\infty} A_{2^n}\big) = \lim_{n\to\infty} \mathbb{P}(A_{2^n}) \le \frac{1}{\lambda^p} \mathbb{E}(|M_T|^p).$$

□

4.4 Continuity of paths

We shall use the martingale inequality of Theorem 4.4 in Section 4.3 to prove the existence of a continuous version for our stochastic integrals. Two stochastic processes I_t and J_t are called versions of each other if $I_t(\omega)$ and $J_t(\omega)$ are equal for all $t \in [0,T]$ and almost all $\omega \in \Omega$. (recall Definition 2.3 in Section 2.3).

Theorem 4.5 *Let $f \in \mathcal{L}^2(B, [0,T])$. Let*

$$I_t(\omega) = \int_0^t f(s,\omega) dB_s(\omega).$$

Then there exists a version J_t of I_t such that $t \mapsto J_t(\omega)$ is continuous for almost all $\omega \in \Omega$.

Proof. The point of the proof is to turn continuity in $L^2(\Omega, \mathbb{P})$ into continuity of paths. This requires several estimates.

Let $\phi_n \in \mathcal{L}^2(B, [0,T])$ be an approximation of f by step functions. Put

$$I_n(t, \omega) = \int_0^t \phi_n(s, \omega) dB_s(\omega).$$

By Lemma 4.3 in Section 4.3, I_n is a pathwise continuous martingale for all n. The same holds for the differences $I_n - I_m$. Therefore, by Doob's martingale inequality (with $p = 2$ and $\lambda = \varepsilon$) and the Itô-isometry, we have

$$\begin{aligned}\mathbb{P}\left[\sup_{0\le t\le T} |I_n(t) - I_m(t)| \ge \varepsilon\right] &\le \frac{1}{\varepsilon^2}\mathbb{E}\big((I_n(T) - I_m(T))^2\big) \\ &= \frac{1}{\varepsilon^2}\int_0^T \mathbb{E}\big((\phi_n(t) - \phi_m(t))^2\big) dt \\ &= \frac{1}{\varepsilon^2}\|\phi_n - \phi_m\|^2_{\mathcal{L}^2(B,[0,T])},\end{aligned}$$

which tends to 0 as $n, m \to \infty$ because ϕ_n is a Cauchy sequence. We can therefore choose an increasing sequence $n_1, n_2, n_3, \ldots$ of natural numbers such that

$$\mathbb{P}\left[\sup_{0\le t\le T} |I_{n_{k+1}}(t) - I_{n_k}(t)| > 2^{-k}\right] \le 2^{-k}.$$

By the first Borel-Cantelli lemma it follows that

$$\mathbb{P}\left[\sup_{0\le t\le T} |I_{n_{k+1}}(t) - I_{n_k}(t)| > 2^{-k} \text{ for infinitely many } k\right] = 0.$$

Hence for almost all ω there exists $K(\omega)$ such that for all $k \ge K(\omega)$,

$$\sup_{0\le t\le T} |I_{n_{k+1}}(t, \omega) - I_{n_k}(t, \omega)| \le 2^{-k},$$

so that for $l > k > K(\omega)$,

$$\sup_{0\le t\le T} |I_{n_l}(t, \omega) - I_{n_k}(t, \omega)| \le \sum_{j=k}^{l-1} 2^{-j} \le 2^{-(k-1)}.$$

This implies that $t \mapsto I_{n_k}(t, \omega)$ as $k \to \infty$ converges uniformly to some function $t \mapsto J(t, \omega)$, which must therefore be continuous. It remains to show that $J(t, \omega)$ is a version of $I(t, \omega)$. This can be done by Fatou's lemma, namely for all $t \in [0, T]$:

$$\begin{aligned}\int_\Omega |J(t, \omega) - I(t, \omega)|^2 \mathbb{P}(d\omega) &= \int_\Omega \liminf_{k\to\infty} |I_{n_k}(t, \omega) - I(t, \omega)|^2 \mathbb{P}(d\omega) \\ &\le \liminf_{k\to\infty} \int_\Omega |I_{n_k}(t, \omega) - I(t, \omega)|^2 \mathbb{P}(d\omega) = 0.\end{aligned}$$

□

From now on we shall always take $\int_0^t f(s, \omega) dB_s$ to mean a t-continuous version of the integral.

We have completed our construction of stochastic integrals. In Sections 5–8 we shall investigate their main properties.

5 Stochastic integrals and the Itô-formula

In this chapter we shall treat the Itô-formula, a stochastic chain rule that is of great help in the formal manipulation of stochastic integrals.

We say that a process X_t is a *stochastic integral* if there exist (square-integrable adapted) processes $U_t, V_t \in \mathcal{L}^2(B, [0,T])$ such that, for all $t \in [0,T]$,

$$X_t = X_0 + \int_0^t U_s ds + \int_0^t V_s dB_s. \tag{5.1}$$

The first integral on the r.h.s. is of finite variation, being pathwise differentiable almost everywhere. The second integral is an Itô-integral and therefore a martingale. A decomposition of a process into a martingale and a process of finite variation is called a *Doob-Meyer decomposition*. Processes in $\mathcal{L}^2(B, [0,T])$ that have such a decomposition are called "semi-martingales". Equation (5.1) is conveniently rewritten in differential form:

$$dX_t = U_t dt + V_t dB_t. \tag{5.2}$$

Example 5.1.¡ In Section 4.2 it was shown that the process B_t^2 satisfies the equation

$$d(B_t^2) = dt + 2B_t dB_t. \tag{5.3}$$

5.1 The one-dimensional Itô-formula

Relation (5.3) is an instance of a general chain rule for functions of stochastic integrals that can be stated as follows: "the differential must be expanded to second order and every occurrence of $(dB_t)^2$ must be replaced by dt". Here is the precise rule.

Theorem 5.1 (Itô-formula). *Let X_t be a stochastic integral. Let $g\colon [0,\infty) \times \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable. Then the process $Y_t := g(t, X_t)$ satisfies*

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \tfrac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t)\,(dX_t)^2\,, \tag{5.4}$$

where $(dX_t)^2$ is to be evaluated according to the multiplication table:

	dt	dB_t
dt	0	0
dB_t	0	dt

i.e., with the Doob-Meyer decomposition (5.1):

$$\begin{aligned} dX_t &= U_t dt + V_t dB_t \\ (dX_t)^2 &= V_t^2 dt. \end{aligned}$$

In terms of the explicit form (5.2) of X_t, we can write (5.4) as

$$dY_t = U_t' dt + V_t' dB_t$$

with

$$U_t' = \frac{\partial g}{\partial t}(t, X_t) + \frac{\partial g}{\partial x}(t, X_t)\, U_t + \tfrac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t)\, V_t^2$$
$$V_t' = \frac{\partial g}{\partial x}(t, X_t)\, V_t,$$

which in turn stands for

$$Y_T = Y_0 + \int_0^T U_s' ds + \int_0^T V_s' dB_s.$$

The third term in the r.h.s. of (5.4) is called "the Itô correction".

We shall prove Theorem 5.1 via the following extension of Lemma 2.3 in Section 2.4.

Lemma 5.2 *If A_t is a process in $\mathcal{L}^2(B, [0, T])$, then*

$$\sum_{j=0}^{n-1} A_{t_j} (\Delta B_j)^2 \longrightarrow \int_0^T A_t dt \qquad \text{in } L^2(\Omega, \mathbb{P}) \text{ as } n \longrightarrow \infty.$$

Proof. We leave this as an exercise. □

We now give the proof of Theorem 5.1. It will be a bit sketchy, but the details are easily filled in.

Proof. We shall use the by now standard notation

$$\Delta t_j = \Delta t_j^{(n)} = t_{j+1}^{(n)} - t_j^{(n)}, \quad \text{and} \quad \Delta B_j = \Delta B_j^{(n)} = B_{t_{j+1}^{(n)}} - B_{t_j^{(n)}},$$

where the points $(t_j^{(n)})_{j=0}^n$ with

$$0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = T \quad \text{for} \quad n = 1, 2, 3, \ldots$$

form a sequence of partitions of $[0, T]$ whose mazes $\delta_n = \max_{0 \le j < n} \Delta t_j^{(n)}$ tend to zero as $n \longrightarrow \infty$. We shall generally write X_j for X_{t_j}.

1. First, we may assume that g, $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$, $\frac{\partial^2 g}{\partial t^2}$, $\frac{\partial^2 g}{\partial x^2}$ and $\frac{\partial^2 g}{\partial x \partial t}$ are all bounded. If they are not, then we stop the process as soon as the absolute value of one of them reaches the value N, and afterwards take the limit $N \longrightarrow \infty$.

2. Next, using Taylor's theorem we obtain

$$\begin{aligned} g(T,X_T)-g(0,X_0) &= \sum_j \left(g(t_{j+1},X_{j+1})-g(t_j,X_j)\right) \\ &= \sum_j \frac{\partial g}{\partial t}(t_j,X_j)\Delta t_j + \sum_j \frac{\partial g}{\partial x}(t_j,X_j)\Delta X_j \\ &\quad + \tfrac{1}{2}\sum_j \frac{\partial^2 g}{\partial t^2}(t_j,X_j)\,(\Delta t_j)^2 + \sum_j \frac{\partial^2 g}{\partial t\partial x}(t_j,X_j)\Delta t_j \Delta X_j \\ &\quad + \tfrac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}(t_j,X_j)\,(\Delta X_j)^2 + \sum_j \varepsilon_j, \end{aligned}$$

where $\varepsilon_j = \mathrm{o}\left(|\Delta t_j|^2 + |\Delta X_j|^2\right)$ for all j.

3. The first two terms converge because $\delta_n \to 0$:

$$\begin{aligned} \sum_j \frac{\partial g}{\partial t}(t_j,X_j)\Delta t_j &\longrightarrow \int_0^T \frac{\partial g}{\partial t}\,(t,X_t)\,dt, \\ \sum_j \frac{\partial g}{\partial x}(t_j,X_j)\Delta X_j &= \sum_j \frac{\partial g}{\partial x}(t_j,X_j)\,(U_j\Delta t_j + V_j \Delta B_j) + o(1) \\ &\longrightarrow \int_0^T \frac{\partial g}{\partial x}\,(t,X_t)\,U_t dt + \int_0^T \frac{\partial g}{\partial x}\,(t,X_t)\,V_t dB_t. \end{aligned}$$

4. The third and the fourth term tend to zero. For instance, if in the fourth term we substitute $\Delta X_j = U_j \Delta t_j + V_j \Delta B_j$, then a term

$$\sum_j \frac{\partial^2 g}{\partial t \partial x}(t_j,X_j) V_j \Delta t_j \Delta B_j =: \sum_j c_j \Delta t_j \Delta B_j \tag{5.5}$$

arises. But, because c_j is $\mathcal{F}_{t_j}$-measurable and $|c_j| \le M$ for all j, it follows that (5.5) tends to zero because

$$\mathbb{E}\Big(\big(\sum_j c_j \Delta t_j \Delta B_j\big)^2\Big) = \sum_j \mathbb{E}(c_j^2)(\Delta t_j)^3 \longrightarrow 0.$$

5. The fifth term again converges:

$$\begin{aligned} \tfrac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}\,(\Delta X_j)^2 &= \tfrac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}(t_j,X_j)\,(U_j \Delta t_j + V_j \Delta B_j)^2 + o(1) \\ &= \tfrac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}(t_j,X_j)\,\left(U_j^2(\Delta t_j)^2 + 2U_jV_j\Delta t_j \Delta B_j + V_j^2(\Delta B_j)^2\right) + o(1) \\ &\longrightarrow \tfrac{1}{2}\int_0^T \frac{\partial^2 g}{\partial x^2}(t_j,X_j) V_t^2 dt \end{aligned}$$

by Lemma 5.2 (recall the multiplication table in Theorem 5.1). □

5.2 Some examples

Example 5.2.¡ We can now generalise (5.3) as follows:

$$d\left(B_t^n\right) = nB_t^{n-1}\, dB_t + \frac{1}{2}n(n-1)B_t^{n-2}\, dt\,, \tag{5.6}$$

(use Theorem 5.1 with $g(t,x) = x^n$, $U_t \equiv 0$.)

Example 5.3.¡ More generally, we have for every twice continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$,

$$df(B_t) = f'(B_t)\, dB_t + \frac{1}{2}f''(B_t)\, dt\,. \tag{5.7}$$

Example 5.4.¡ With the help of Itô's formula (5.6) it is possible to quickly calculate an integral like $\int_0^T B_t^2 dB_t$, in much the same way as ordinary integrals. We make a guess for the indefinite integral, calculate its derivative, and where needed we apply a correction. In the case at hand we would guess the integral to be something like B_t^3, so we calculate $V_t \equiv 1$):

$$d(B_t^3) = 3B_t^2 dB_t + 3B_t(dB_t)^2 = 3B_t^2 dB_t + 3B_t dt$$

$$\Longrightarrow B_t^2 dB_t = \tfrac{1}{3}d\big(B_t^3\big) - B_t dt$$

$$\Longrightarrow \int_0^T B_t^2 dB_t = \tfrac{1}{3}B_T^3 - \int_0^T B_t dt.$$

Example 5.5.¡ In the same way it is found that

$$\int_0^T \sin B_t\, dB_t = 1 - \cos B_T - \frac{1}{2}\int_0^T \cos B_t\, dt\,.$$

Example 5.6.¡ Let f be differentiable. We can rewrite $N_f = \int_0^T f(t)\, dB_t$ as follows, (use Theorem 5.1 with $g(t,x) = f(t)x$, $U_t \equiv 0$, $V_t \equiv 1$):

$$d\left(f(t)B_t\right) = d(f(t))B_t + f(t)dB_t$$

$$\Longrightarrow \int_0^T f(t)dB_t = f(T)B_T - f(0)B_0 - \int_0^T f'(t)B_t dt.$$

This is a partial integration formula for the integration of functions with respect to Brownian motion.

Example 5.7.¡ Let us solve the stochastic differential equation

$$dX_t = \beta X_t\, dB_t,$$

which is a special case of the growth equation in Example 1 in Section 1. We try $Y_t = \exp(\beta B_t)$, and obtain with the help of the Itô-formula that

$$dY_t = \beta Y_t\, dB_t + \tfrac{1}{2}\beta^2 Y_t\, dt.$$

This is obviously growing too fast: the second term in the r.h.s., which is the Itô-correction, must be compensated. We therefore try next $X_t = \exp(-\alpha t) Y_t$, finding

$$\begin{aligned} dX_t &= -\alpha e^{-\alpha t} Y_t dt + e^{-\alpha t} dY_t \\ &= -\alpha X_t\, dt + (\beta X_t dB_t + \frac{1}{2}\beta^2 X_t dt) \end{aligned}$$

The dt terms cancel if $\alpha = \frac{1}{2}\beta^2$ and we find the solution

$$X_t = \mathrm{e}^{\beta B_t - \frac{1}{2}\beta^2 t}.$$

This process is called the *exponential martingale.*

5.3 The multi-dimensional Itô-formula

Theorem 5.3 *Let $B(t,\omega) = (B_1(t,\omega), B_2(t,\omega), \ldots, B_m(t,\omega))$ be Brownian motion in $\mathbb{R}^m$, consisting of m independent copies of Brownian motion. Let $X(t,\omega) = (X_1(t,\omega), \ldots, X_1(t,\omega))$ be the stochastic integral given by*

$$dX_i(t) = U_i(t)dt + \sum_{j=1}^{m} V_{ij}(t) dB_j(t) \qquad (i = 1, \ldots, n), \tag{5.8}$$

for some processes $U_i(t)$ and $V_{ij}(t)$ in $\mathcal{L}^2(B, [0,T])$. Abbreviate (5.8) in the vector notation

$$dX = U dt + V dB.$$

Let $g\colon\ [0,T] \times \mathbb{R}^n \longrightarrow \mathbb{R}^p$ be a C^2-function. Then the process $Y(t,\omega) := g(t, X_1(t), \ldots, X_n(t))$ satisfies the stochastic differential equation

$$dY_i(t) = \frac{\partial g_i}{\partial t}(t, X_t)\, dt + \sum_{j=1}^{n} \frac{\partial g_i}{\partial x_j}(t, X_t)\, dX_j(t) + \tfrac{1}{2} \sum_{j,k=1}^{n} \frac{\partial^2 g_i}{\partial x_j \partial x_k}(t, X_t)\, dX_j(t) dX_k(t) \tag{5.9}$$

$(i = 1, \ldots, p)$,

where the product $dX_j dX_k$ has to be evaluated according to the rules

$$\begin{aligned} dB_j\, dB_k &= \delta_{jk}\, dt \\ dB_j\, dt &= (dt)^2 = 0. \end{aligned}$$

Equation 5.9 is the multi-dimensional version of Itô's formula, and can be proved in the same way as its one-dimensional counterpart. Be careful to keep track of all the indices. The easiest case is $m = n$, $p = 1$.

Example 5.8.¡ (Bessel process). Let $R_t(\omega) = \|B_t(\omega)\|$, where B_t is m-dimensional Brownian motion and $\|\cdot\|$ is the Euclidean norm. Apply the Itô-formula to the function $r\colon \mathbb{R}^m \longrightarrow \mathbb{R}_+ : x \mapsto \|x\|$. We compute $\frac{\partial r}{\partial x_i} = \frac{x_i}{r}$ and $\frac{\partial^2 r}{\partial {x_i}^2} = \frac{1}{r} - \frac{x_i^2}{r^3}$. So we find

$$dR = \sum_{j=1}^{m} \frac{B_j}{R} dB_j + \tfrac{1}{2} \sum_{j=1}^{m} \left(\frac{1}{R} - \frac{B_j^2}{R^3} \right) dt = \sum_{j=1}^{m} \frac{B_j}{R} dB_j + \frac{m-1}{2R} dt.$$

For notational convenience we have dropped here the arguments t and ω.

The next theorem gives a way to construct martingales out of Brownian motion. A function f on $\mathbb{R}^m$ is called *harmonic* if $\Delta f = 0$, where Δ denotes the Laplace operator $\sum_{i=1}^m \frac{\partial^2}{\partial x_i^2}$.

Theorem 5.4 *If $f\colon \mathbb{R}^m \longrightarrow \mathbb{R}$ is harmonic, then $f(B_t)$ is a martingale.*

Proof. Write out

$$df(B_t) = \sum_{i=1}^{m} \frac{\partial f}{\partial x_i} (B(t))\, dB_i(t) + \tfrac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j} (B(t))\, dB_i(t) dB_j(t).$$

The second part in the r.h.s. is zero because $dB_i(t)dB_j(t) = \delta_{ij} dt$ and $\Delta f = 0$. Integration yields

$$f(B_t) - f(B_0) = \sum_{i=1}^{m} \int_0^t \frac{\partial f}{\partial x_i} (B(t))\, dB_i(t).$$

This is an Itô-integral and hence a martingale. □

We can understand Theorem 5.4 intuitively as follows. A harmonic function f has the property that its value in a point x is the average over its values on any sphere around x. This property, together with the fact that multi-dimensional Brownian motion is isotropic, explains why $f(B_t)$ is a "fair game".

The following technical extension of Itô's formula (5.7) will be useful later on.

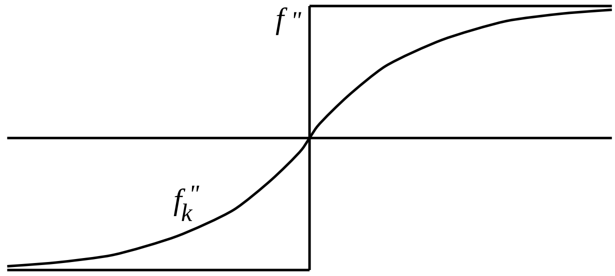


Figure 3: The graph of f_k'' approximating f'' having a jump.

Lemma 5.5 *Itô's formula for a function of Brownian motion*

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)\, dB_s + \tfrac{1}{2} \int_0^t f''(B_s)\, ds \tag{5.10}$$

still holds if $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ *is* C^1 *everywhere and* C^2 *outside a finite set* $\{ z_1, \ldots, z_N \}$, *with* f'' *bounded on some neighbourhood of this set.*

Proof. Take $f_k \in \mathrm{C}^2(\mathbb{R})$ such that $f_k \longrightarrow f$ and $f_k' \longrightarrow f'$ as $k \to \infty$, both uniformly, and such that for $x \notin \{ z_1, \ldots, z_N \}$:

$$\begin{cases} f_k''(x) \longrightarrow f''(x) \text{ as } k \to \infty \\ |f_k''(x)| \leq M \text{ in a neighbourhood of } \{ z_1, \ldots, z_N \}. \end{cases}$$

(Fig. 3 shows the graph of f_k'' approximating some f'' with a jump.) For f_k we have the Itô formula

$$f_k(B_t) = f_k(B_0) + \int_0^t f_k'(B_s) dB_s + \tfrac{1}{2} \int_0^t f_k''(B_s) ds.$$

In the limit as $k \longrightarrow \infty$ this equality tends, term by term in $L^2(\Omega, \mathbb{P})$, to (5.10). Indeed, the distances $\|f_k(B_t) - f(B_t)\|$ and $\|f_k(B_0) - f(B_0)\|$ tend to 0 by the uniformity of the convergence $f_k \to f$, the norm difference $\| \int_0^t f_k'(B_s)\, dB_s - \int_0^t f'(B_s)\, dB_s\|$ tends to 0 by the uniformity of the convergence $f_k' \to f'$ plus Itô's isometry, and the difference

$$\mathbb{E}\left(\int_0^t f_k''(B_s)\, ds - \int_0^t f''(B_s)\, ds \right)^2 \leq t \int_o^t \int_\Omega \big(f_k''(B_s(\omega)) - f''(B_s(\omega))\big)^2 \mathbb{P}(d\omega)\, ds$$

tends to 0 by dominated convergence combined with the fact that

$$\mathbb{P} \otimes \lambda\Big(\Big\{(\omega, s) \in \Omega \times [0,t] \,:\, B_s(\omega) \in \{z_1, \ldots, z_N\}\Big\}\Big) = 0.$$

□

5.4 Local times of Brownian motion

As an application of Lemma 5.5 we shall prove Tanaka's formula for the local times of Brownian motion.

Theorem 5.6 (Tanaka) *Let* $t \mapsto B_t$ *be a one-dimensional Brownian motion. Let* λ *denote the Lebesgue measure on* $[0,T]$. *Then the limit*

$$L_t := \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \lambda\Big(\{ s \in [0,t] \,:\, B_s \in (-\varepsilon, \varepsilon) \}\Big)$$

exists in $L^2(\Omega, \mathbb{P})$ *and is equal to*

$$L_t = |B_t| - |B_0| - \int_0^t \operatorname{sgn}(B_s) dB_s.$$

(Think of L_t as the density per unit length of the total time spent close to the origin up to time t.)

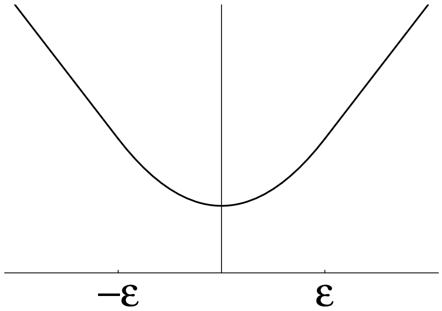


Figure 4: The function $g_\varepsilon(t)$.

Proof. For $\varepsilon > 0$, consider the function

$$g_\varepsilon(x) = \begin{cases} |x| & \text{if } |x| \geq \varepsilon \\ \frac{1}{2}\left(\varepsilon + x^2/\varepsilon\right) & \text{if } |x| < \varepsilon, \end{cases}$$

as shown in Fig. 4. Then g_ε is C^2, except in the points $\{-\varepsilon, \varepsilon\}$, and it is C^1 everywhere on $\mathbb{R}$. Apply Lemma 5.5 to get

$$\tfrac{1}{2}\int_0^t g_\varepsilon''(B_s)ds = g_\varepsilon(B_t) - g_\varepsilon(B_0) - \int_0^t g_\varepsilon'(B_s)dB_s, \tag{5.11}$$

where $g_\varepsilon'(x)$ and $g_\varepsilon''(x)$ are given by

$$g_\varepsilon'(x) = \begin{cases} \operatorname{sgn}(x) & \text{if} \quad |x| \geq \varepsilon \\ x/\varepsilon & \text{if} \quad |x| < \varepsilon \end{cases}$$

and

$$g_\varepsilon''(x) = \tfrac{1}{\varepsilon}\mathbb{1}_{(-\varepsilon,\varepsilon)}(x) \qquad \text{if } x \notin \{-\varepsilon, \varepsilon\}.$$

Now, the limit as $\varepsilon \downarrow 0$ of the l.h.s. of (5.11) is precisely L_t. Moreover, we trivially have $g_\varepsilon(B_t) \to |B_t|$ and $g_\varepsilon(B_0) \to |B_0|$ as $\varepsilon \downarrow 0$. Hence it suffices to prove that the integral in the r.h.s. of (5.11) converges to the appropriate limit:

$$\int_0^t \left(g_\varepsilon'(B_s) - \operatorname{sgn}(B_s)\right) dB_s \longrightarrow 0 \quad \text{in} \quad L^2(\Omega, \mathbb{P}).$$

To see why the latter is true, estimate

$$\begin{aligned} \left\| \int_0^t \left(g_\varepsilon'(B_s) - \operatorname{sgn}(B_s)\right) dB_s \right\|^2 &= \left\| \int_0^t \mathbb{1}_{(-\varepsilon,\varepsilon)}(B_s)\left(\frac{1}{\varepsilon}B_s - \operatorname{sgn}(B_s)\right) dB_s \right\|^2 \\ &= \mathbb{E}\left(\int_0^t \mathbb{1}_{(-\varepsilon,\varepsilon)}(B_s)\left(\frac{1}{\varepsilon}B_s - \operatorname{sgn}(B_s)\right)^2 ds \right) \\ &\leq \int_0^t \mathbb{P}\left(B_s \in (-\varepsilon, \varepsilon)\right) ds \\ &\longrightarrow 0 \quad \text{as} \quad \varepsilon \downarrow 0, \end{aligned}$$

where in the second equality we use the Itô-isometry, in the inequality we use that $|B_s| < \varepsilon$ implies that $|\varepsilon^{-1}B_s - \text{sgn}(B_s)| \leq 1$, and the last statement holds because the Gaussian random variable B_s ($s > 0$) has finite density. It follows that L_t exists and can be expressed as in the statement of the theorem. □

Note that for smooth functions f,

$$|f(t)| - |f(0)| - \int_0^t \text{sgn}\,(f(s))\, f'(s)ds = 0$$

because

$$\frac{d}{dt}|f(t)| = \text{sgn}\,(f(t))\, f'(t) \qquad (f(t) \neq 0).$$

Thus, the local time is an Itô-correction to this relation, caused by the fact that $d|B_t| \neq \text{sgn}\,(B_t)\, dB_t$: if B_t passes the origin during the time interval Δt_j, then $|B_{t_{j+1}} - B_{t_j}|$ need not be equal to $\text{sgn}(B_{t_j})\Delta B_{t_j}$. The difference is a measure of the time spent close to the origin.

The existence of the local times of Brownian motion was proved by Lévy in the 1930's using hard estimates. The above approach is shorter and more elegant. What is described above is the local time at the origin: $L_t = L_t(0)$. In a completely analogous way one can prove the existence of the local time $L_t(x)$ at any site $x \in \mathbb{R}$.

The process plays a key role in many applications associated with Brownian motion. For instance, it is used to define path measures that model the behaviour of polymer chains. Let $\mathbb{P}_T$ denote the Wiener measure on the time interval $[0, T]$. Fix a number $\beta > 0$, and define a new measure $\mathbb{Q}_T^\beta$ on path space by setting

$$\frac{d\mathbb{Q}_T^\beta}{d\mathbb{P}_T}(\cdot) = \frac{1}{Z_T^\beta} \exp\left[-\beta \int_{\mathbb{R}} L_T^2(x)(\cdot)\, dx\right] ,$$

where Z_T^β is the normalising constant. What the measure $\mathbb{Q}_T^\beta$ does is reward paths that spread themselves out compared to Brownian motion. We may think of $\mathbb{P}_T$ as modeling the erratic spatial distribution of the polymer and of the exponential weight factor as modeling its stiffness or self-repellence. The parameter β is the strength of the self-repellence. It is known that

$$\lim_{T\to\infty} \mathbb{Q}_T^\beta \left(\frac{|X_T|}{T} \in [\vartheta(\beta) - \varepsilon, \vartheta(\beta) + \varepsilon]\right) = 1 \quad \forall \varepsilon > 0$$

for some constant $\vartheta(\beta)$, called the *asymptotic speed* of the polymer, given by

$$\vartheta(\beta) = \text{Cst} \cdot \beta^{\frac{1}{3}}.$$

(See R. van der Hofstad, F. den Hollander and W. König (1997).) The Brownian motion corresponds to $\beta = 0$ and has zero speed.

6 Stochastic differential equations

A *stochastic differential equation* for a process $X(t)$ with values in $\mathbb{R}^n$ is an equation of the form

$$dX_i(t) = b_i(t, X_t)\, dt + \sum_{j=1}^{m} \sigma_{ij}(t, X_t)\, dB_j(t) \qquad (i = 1, \cdots, n). \tag{6.1}$$

Here, $B(t) = (B_1(t), B_2(t), \cdots, B_m(t))$ is an m-dimensional Brownian motion, i.e., an m-tuple of independent Brownian motions on $\mathbb{R}$. The functions b_i and σ_{ij} from $\mathbb{R} \times \mathbb{R}^n$ to $\mathbb{R}$ with $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$ form a field b of n-vectors and a field σ of $n \times m$-matrices. A process $X \in \mathcal{L}^2(B, [0, T])$ for which (6.1) holds is called a *strong solution* of the equation. In more pictorial language, such a solution is called an *Itô-diffusion* with *drift* b and *diffusion matrix* $\sigma\sigma^*$.

In this section we shall formulate a result on the existence and the uniqueness of Itô-diffusions. It will be convenient to employ the following notation for the norms on vectors and matrices:

$$\|x\|^2 := \sum_{i=1}^{n} x_i^2 \qquad (x \in \mathbb{R}^n); \qquad \|\sigma\|^2 := \sum_{i=1}^{n} \sum_{j=1}^{m} \sigma_{ij}^2 = tr(\sigma\sigma^*) \qquad (\sigma \in \mathbb{R}^{n \times m}).$$

Also, we would like to take into account an initial condition $X(0) = Z$, where Z is an $\mathbb{R}^n$-valued random variable independent of the Brownian motion. All in all, we enlarge our probability space and our space of adapted processes as follows. We choose a probability measure μ on $\mathbb{R}^n$ and put

$$\begin{aligned}
\Omega' &:= \mathbb{R}^n \times \Omega^m\,; \\
\mathcal{F}'_t &:= \mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}_t^{\otimes m} \quad (t \in [0, T]); \\
\mathbb{P}' &:= \mu \otimes \mathbb{P}^{\otimes m}; \\
Z &: \Omega' \to \mathbb{R}^n : (z, \omega) \mapsto z; \\
\mathcal{L}^2(B, [0, T])' &:= \{X \in L^2(\mathbb{R}^n, \mu) \otimes L^2(\Omega, \mathcal{F}_T, \mathbb{P})^{\otimes m} \otimes L^2[0, T] \otimes \mathbb{R}^n \mid \\
&\qquad \omega \mapsto X_{t,i}^x(\omega) \text{ is } \mathcal{F}'_t\text{-measurable}\}.
\end{aligned}$$

This change of notation being understood, we shall drop the primes again.

6.1 Strong solutions

Theorem 6.1 *Fix $T > 0$. Let $b\colon\ [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma\colon\ [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$ be measurable functions, satisfying the growth conditions*

$$\|b\,(t, x)\| \vee \|\sigma(t, x)\| \leq C\,(1 + \|x\|) \quad (t \in [0, T],\ x \in \mathbb{R}^n)$$

as well as the Lipschitz conditions

$$\|b(t, x) - b(t, y)\| \vee \|\sigma(t, x) - \sigma(t, y)\| \leq D\,\|x - y\| \quad (t \in [0, T],\ x, y \in \mathbb{R}^n),$$

for some positive constants C and D. Let $(B_t)_{t\in[0,T]}$ be m-dimensional Brownian motion and let μ be a probability measure on $\mathbb{R}^n$ with $\int_{\mathbb{R}^n}\|x\|^2\mu(dx)<\infty$. Then the stochastic differential equation (6.1) has a unique continuous adapted solution $X(t)$ given that the law of $X(0)$ is equal to μ.

Proof. The proof comes in three parts.

1. Uniqueness. Suppose $X, Y \in \mathcal{L}^2(B,[0,T])$ are solutions of (6.1) with continuous paths. (Continuity will be proved in part 3). Put

$$\Delta b_i(t) := b_i(t,X_t) - b_i(t,Y_t)$$
$$\Delta\sigma_{ij}(t) := \sigma_{ij}(t,X_t) - \sigma_{ij}(t,Y_t).$$

Applying the inequality $(a+b)^2 \le 2(a^2+b^2)$ for real numbers a and b, the independence of the components of $B(t)$, the Cauchy-Schwarz inequality $(\int_0^t g(s)ds)^2 \le t\int_0^t g(s)^2 ds$ for an L^2-function g, the multi-dimensional Itô-isometry and finally the Lipschitz condition, we find

$$\begin{aligned}
\mathbb{E}\left(\|X_t - Y_t\|^2\right) &:= \sum_{i=1}^n \mathbb{E}\big((X_i(t)-Y_i(t))^2\big) \\
&= \sum_{i=1}^n \mathbb{E}\left(\left(\int_0^t \Delta b_i(s)ds + \sum_{j=1}^m \int_0^t \Delta\sigma_{ij}(s)dB_j(s)\right)^2\right) \\
&\le 2\sum_{i=1}^n \mathbb{E}\left(\left(\int_0^t \Delta b_i(s)ds\right)^2 + \left(\sum_{j=1}^m \int_0^t \Delta\sigma_{ij}(s)dB_j(s)\right)^2\right) \\
&= 2\sum_{i=1}^n \mathbb{E}\left(\left(\int_0^t \Delta b_i(s)ds\right)^2\right) + 2\sum_{i=1}^n \mathbb{E}\left(\sum_{j=1}^m \int_0^t (\Delta\sigma_{ij}(s))^2 ds\right) \\
&\le 2t\int_0^t \mathbb{E}\left(\sum_{i=1}^n \Delta b_i(s)^2\right)ds + 2\int_0^t \mathbb{E}\left(\sum_{i=1}^n\sum_{j=1}^m \Delta\sigma_{ij}(s)^2\right)ds \\
&= 2t\int_0^t \mathbb{E}(\|\Delta b(s)\|^2)ds + 2\int_0^t \mathbb{E}(\|\Delta\sigma(s)\|^2)ds \\
&\le 2(t+1)D^2\int_0^t \mathbb{E}\|X_s - Y_s\|^2 ds.
\end{aligned}$$

So the function $f\colon t \mapsto \mathbb{E}\left(\|X_t - Y_t\|^2\right)$ satisfies the integral inequality

$$0 \le f(t) \le A\int_0^t f(s)ds \qquad (t\in[0,T]) \tag{6.2}$$

for the constant $A = 2D^2(T+1)$.

Lemma 6.2 (Gronwall's lemma) *Inequality (6.2) implies that $f = 0$.*

Proof. Put $F(t) = \int_0^t f(s)ds$. Then $F(t)$ is C^1 and $F'(t) = f(t) \leq AF(t)$. Therefore

$$\frac{d}{dt}\left(\mathrm{e}^{-tA}F(t)\right) = \mathrm{e}^{-tA}\left(f(t) - AF(t)\right) \leq 0.$$

Since $F(0) = 0$, it follows that $\mathrm{e}^{-tA}F(t) \leq 0$ implying $F(t) \leq 0$. So we have $0 \leq f(t) \leq AF(t) \leq 0$ and hence $f(t) = 0$. □

Thus, we have $\mathbb{E}\left(\|X_t - Y_t\|^2\right) = 0$ for all $t \in [0,T]$. In particular,

$$\forall t \in [0,T] \cap \mathbb{Q}: \qquad X_t(\omega) = Y_t(\omega) \text{ for almost all } \omega.$$

Now let

$$N := \{\, \omega \in \Omega \mid \exists t \in [0,T] \cap \mathbb{Q} : X_t(\omega) \neq Y_t(\omega) \,\}.$$

Then N is a countable union of null sets, so $\mathbb{P}(N) = 0$. For $\Omega \setminus N$ we have for all $t \in [0,T] \cap \mathbb{Q}$,

$$X_t(\omega) = Y_t(\omega)\ ,$$

and since X_t and Y_t have continuous paths we conclude that for almost all ω the equality extends to all $t \in [0,T]$. This completes the proof of uniqueness.

2. Existence. We shall find a solution of (6.1) by iterating the map

$$\mathcal{L}^2(B,[0,T]) \to \mathcal{L}^2(B,[0,T])\colon X \to \tilde{X}$$

defined by

$$\tilde{X}_t = Z + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dB_s.$$

Let us start with the constant process $X_t^0 := Z$, and define recursively

$$X_t^{(k+1)} := \tilde{X}_t^{(k)} \qquad (k \geq 0).$$

The calculation in part 1 can be used to conclude that

$$\mathbb{E}\left(\left\|\tilde{X}_t - \tilde{Y}_t\right\|^2\right) \leq A\int_0^t \mathbb{E}\left(\|X_s - Y_s\|^2\right) ds \quad \text{for any} \quad X_s, Y_s \in \mathcal{L}^2(B).$$

Iteration of this result yields, for $k \geq 1$ and the choice $\tilde{X}_t = \tilde{X}_t^{(k)}$ and $\tilde{Y}_t = \tilde{X}_t^{(k-1)}$,

$$\begin{aligned}
\mathbb{E}\left(\left\|X_t^{(k+1)} - X_t^{(k)}\right\|^2\right) &= \mathbb{E}\left(\left\|\tilde{X}_t^{(k)} - \tilde{X}_t^{(k-1)}\right\|^2\right) \\
&\leq A\int_0^t \mathbb{E}\left(\left\|X_s^{(k)} - X_s^{(k-1)}\right\|^2\right) ds \\
&\leq A^k \int_0^t ds_{k-1} \int_0^{s_{k-1}} ds_{k-2} \cdots \int_0^{s_1} ds_0\ \mathbb{E}\left(\left\|X_{s_0}^{(1)} - Z\right\|^2\right) \\
&\leq \frac{A^k t^k}{k!} K \qquad (t \in [0,T]),
\end{aligned}$$

where we insert

$$\tilde{X}^{(0)}_{s_0} = Z + \int_0^t b(s,Z)ds + \int_0^t \sigma(s,Z)dB_s$$

and K is given by

$$\begin{aligned} K &:= \sup_{t\in[0,T]} \mathbb{E}\left(\left\|\int_0^t b(s,Z)ds + \int_0^t \sigma(s,Z)dB_s\right\|^2\right) \\ &\leq 2 \sup_{t\in[0,T]} \left\{\mathbb{E}\left(\left\|\int_0^t b(s,Z)ds\right\|^2\right) + \mathbb{E}\left(\left\|\int_0^t \sigma(s,Z)dB_s\right\|^2\right)\right\} \\ &\leq 2T^2C^2\mathbb{E}(1+\|Z\|)^2 + 2TC^2\mathbb{E}(1+\|Z\|)^2 \\ &\leq 2C^2(T^2+T)\mathbb{E}(1+\|Z\|)^2 , \end{aligned}$$

which is finite by the growth condition and the requirement that Z has finite variance. Now let $X = \mathcal{L}^2(B,[0,T]))\text{-}\lim_{k\to\infty} X^{(k)}$. The existence of this limit follows because $\sum_k \sqrt{A^k t^k/k!} < \infty$ (Cauchy sequence). Then

$$\tilde{X} = \left(\mathcal{L}^2(B,[0,T])\text{-}\lim_{k\to\infty} X^{(k)}\right)^{\sim} = \mathcal{L}^2(B,[0,T])\text{-}\lim_{k\to\infty} X^{(k+1)} = X.$$

So X is a solution of (6.1).

3. Continuity and adaptedness. By Theorem 4.5 the paths $t \mapsto X_t(\omega)$ can be assumed to be continuous for almost all $\omega \in \Omega$. The fact that the solution is adapted is immediate from the construction. □

6.2 Weak solutions

What we have shown in Section 6.1 is that there exists a *strong solution* of (6.1), i.e., a solution that is an adapted functional of the Brownian motion. In the literature on stochastic differential equations often a different point of view is taken, namely one where the Brownian motion itself is also considered unknown. This leads to a so-called *weak solution* of stochastic differential equations, i.e., a solution in distribution.

Definition. A *weak solution* of (6.1) is a pair (B_t, X_t), measurable w.r.t. some filtration $(\mathcal{G}_t)_{t\in[0,T]}$ on some probability space $(\Omega,\ \mathcal{G},\ \mathbb{P})$, such that B_t is m-dimensional Brownian motion and such that (6.1) holds.

The key point here is that the filtration need not be $(\mathcal{F}_t)_{t\in[0,T]} = \sigma((B_s)_{s\in[0,t]})$: if it is, then we have a strong solution.

Definition.

1. *Strong uniqueness* is uniqueness of a strong solution.
2. *Weak uniqueness* holds if all weak solutions have the same law.

By the following example we illustrate the difference between these two concepts.

Example 6.1.¡ (Tanaka). This example is related to our example of Brownian local time in Section 5.4. Let B_t be a Brownian motion and define

$$\tilde{B}_t := \int_0^t \operatorname{sgn}(B_s) dB_s.$$

Since $d\tilde{B}_t = \operatorname{sgn}(B_t) dB_t$ we have, by Itô's formula in Lemma 5.4,

$$d(\tilde{B}_t^2) = 2\tilde{B}_t d\tilde{B}_t + (d\tilde{B}_t)^2$$

with $(d\tilde{B}_t)^2 = \operatorname{sgn}(B_t)^2 (dB_t)^2 = (dB_t)^2 = dt$. Hence $\tilde{B}_t$ is a martingale with quadratic variation t (see Section 2.4):

$$\tilde{B}_t^2 - 2\int_0^t \tilde{B}_s d\tilde{B}_s = t.$$

Since $\tilde{B}_t$ satisfies the requirements (**BM1**)–(**BM5**) of Section 1, it must be a Brownian motion.

Turning matters around, B_t is itself a solution of the stochastic differential equation

$$dB_t = \operatorname{sgn}(B_t) d\tilde{B}_t, \tag{6.3}$$

since $B_t = \int_0^t dB_s = \int_0^t \operatorname{sgn}(B_s)^2 dB_s = \int_0^t \operatorname{sgn}(B_s) d\tilde{B}_s$. Because every solution of (6.3) is a Brownian motion, we have weak uniqueness. However, because $-B_t$ obviously is also a solution of (6.3), there are two solutions of (6.3) for a given process $\tilde{B}_t$. In other words, we do not have strong uniqueness.

Taking this argument one step further we find (cf. Theorem 5.6):

$$\tilde{B}(t) = |B(t)| - |B(0)| - L_t,$$

where L_t is the local time at the origin. By its definition, L_t is adapted to the filtration

$$(\mathcal{G}_t)_{t\in[0,T]}$$

generated by $(|B_t|)_{t\in[0,T]}$. Hence so is $\tilde{B}_t$. It follows that $\tilde{\mathcal{F}}_t \subset \mathcal{G}_t$, where

$$(\tilde{\mathcal{F}}_t)_{t\in[0,T]}$$

is the filtration generated by $(\tilde{B}_t)_{t\in[0,T]}$. However, since B_t is not $\mathcal{G}_t$-measurable (its sign is not determined by $\mathcal{G}_t$), it is not $\tilde{\mathcal{F}}_t$-measurable.

Note that (6.3) is an Itô-diffusion with $b(t,x) = 0$ and $\sigma(t,x) = \operatorname{sgn}(x)$. The latter is not Lipschitz, which is why Theorem 6.1 does not apply.

7 Itô-diffusions, generators and semigroups

The goal of this section is to give a functional analytic description of Itô-diffusions that will allow us to bring powerful analytic tools into play.

7.1 Introduction and motivation

An Itô-diffusion X_t^x (with initial condition $X_0^x = x \in \mathbb{R}^n$) is a Markov process in continuous time. If we suppose that the field b of drift vectors and the field σ of diffusion matrices are both time-independent, i.e., X_t^x is the solution starting at x of the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t \tag{7.1}$$

(where b and σ satisfy the growth and the Lipschitz conditions mentioned in Theorem 6.1), then this Markov process is stationary (= time-homogeneous) and can be characterised by its transition probabilities

$$P_t(x, B) := \mathbb{P}\left[X_t^x \in B\right] \quad (t \geq 0),$$

where B runs through the Borel subsets of $\mathbb{R}^n$. These transition probabilities satisfy the one-parameter Chapman-Kolmogorov equation

$$P_{t+s}(x, B) = \int_{y \in \mathbb{R}^n} P_t(x, dy) P_s(y, B). \tag{7.2}$$

An alternative way to characterise an Itô-diffusion is by its action on a sufficiently rich class of functions $f : \mathbb{R}^n \to \mathbb{R}^n$. Namely, if we define S_t by

$$(S_t f)(x) := \mathbb{E}(f(X_t^x)) = \int_{y \in \mathbb{R}^n} P_t(x, dy) f(y), \tag{7.3}$$

then (7.2) leads to

$$S_{t+s} = S_t \circ S_s, \tag{7.4}$$

i.e., the transformations $(S_t)_{t \geq 0}$ form a *one-parameter semigroup*. Such a semigroup is determined by its *generator* A, defined by

$$Af := \lim_{t \downarrow 0} \frac{1}{t}\left(S_t f - f\right), \tag{7.5}$$

which describes the action of the semigroup over infinitesimal time intervals. In what follows we make the above observations more precise and we study the interplay between the diffusion X_t^x, its generator A and its semigroup $(S_t)_{t \geq 0}$.

7.2 Basic properties

Let $t \geq 0$, and for $s \geq t$ let $X_s^{t,x}$ denote the solution of the stochastic differential equation (7.1) with initial condition $X_t^{t,x} = x$.

Lemma 7.1 (Stationarity) *The processes $s \mapsto X_s^{t,x}$ and $s \mapsto X_{s-t}^x$ are equal in law for $s \geq t$.*

Proof. These processes are solutions of the equations

$$dX_s^{t,x} = b(X_s^{t,x})ds + \sigma(X_s^{t,x})d(B_s - B_t)$$
$$dX_{s-t}^x = b(X_{s-t}^x)ds + \sigma(X_{s-t}^x)d(B_{s-t})$$

respectively, for $s \geq t$. Since $B_s - B_t$ and B_{s-t} are Brownian motions on $[t, T]$, both starting at 0, the two processes have the same law by weak uniqueness (which is implied by strong uniqueness obtained in Theorem 6.1). □

Proposition 7.2 (Markov property) *Let X_t^x $(t \geq 0)$ denote the solution of (7.1) with initial condition $X_0^x = x$. Fix $t \geq 0$. Let $C_0(\mathbb{R}^n)$ be the space of all continuous functions $\mathbb{R}^n \to \mathbb{R}$ that tend to 0 at infinity. Then for all $s \geq 0$ the conditional expectation $\mathbb{E}(f(X_{t+s}^x)|\mathcal{F}_t)$ depends on $\omega \in \Omega$ only via X_t^x. In fact,*

$$\mathbb{E}(f(X_{t+s}^x)|\mathcal{F}_t) = (S_s f)(X_t^x).$$

Proof. Fix $t \geq 0$. For $y \in \mathbb{R}^n$ and $s \geq t$, let

$$Y_s(y) := X_{s+t}^{t,y}.$$

Then, according to Lemma 7.1, $Y_s(y)$ and X_s^y have the same law. Hence, for all $f \in C_0(\mathbb{R}^n)$,

$$\mathbb{E}(f(Y_s(y))) = \mathbb{E}(f(X_s^y)) = (S_s f)(y).$$

On the other hand, the random variable $Y_s(y)$ depends on $\omega \in \Omega$ only via the Brownian motion $u \mapsto B_u(\omega) - B_t(\omega)$ for $u \geq t$, so $Y_s(y)$ is stochastically independent of $\mathcal{F}_t$. By strong uniqueness we must have, for $s \geq 0$,

$$Y_s\left(X_t^x\right) = X_{t+s}^x \quad \text{a.s.},$$

since both sides are solutions of the same stochastic differential equation (7.1) for $s \geq 0$ with initial condition $Y_0(X_t^x) = X_t^x$. As X_t^x is $\mathcal{F}_t$-measurable, we now have

$$\mathbb{E}\left(f(X_{t+s}^x|\mathcal{F}_t)\right)(\omega) = \int_{\omega' \in \Omega} f\left(Y_s(X_t^x(\omega))(\omega')\right) \mathbb{P}(d\omega') = (S_s f)\left(X_t^x(\omega)\right),$$

where ω represents the randomness before time t and ω' the randomness after time t. □

Definition. Let $\mathcal{B}$ be a Banach space. A *jointly continuous one-parameter semigroup of contractions on* $\mathcal{B}$ is a family $(S_t)_{t\geq 0}$ of linear operators $S_t : \mathcal{B} \to \mathcal{B}$ satisfying the following requirements:

CS1. $\|S_t f\| \leq \|f\|$ for all $t \geq 0$, $f \in \mathcal{B}$;
CS2. $S_0 f = f$ for all $f \in \mathcal{B}$;
CS3. $S_{t+s} = S_t \circ S_s$ for all $t, s \geq 0$;
CS4. the map $(t, f) \mapsto S_t f$ is jointly continuous $[0, \infty) \times \mathcal{B} \to \mathcal{B}$.

We choose $\mathcal{B} = C_0(\mathbb{R}^n)$. The natural norm on this space is the supremum norm

$$\|f\| := \sup_{x \in \mathbb{R}^n} |f(x)|.$$

For $f \in C_0(\mathbb{R}^n)$ we define

$$(S_t f)(x) := \mathbb{E}(f(X_t^x)),$$

where X_t^x $(t \geq 0)$ denotes the solution of (7.1) with initial condition $X_0^x = x$.

Theorem 7.3 *The operators S_t with $t \geq 0$ form a jointly continuous one-parameter semigroup of contractions $C_0(\mathbb{R}^n) \to C_0(\mathbb{R}^n)$.*

Proof. We prove properties 1–4 in Definition 7.2.

1. Fix $t \geq 0$ and $f \in C_0(\mathbb{R}^n)$. We show that the function $S_t f$ again lies in $C_0(\mathbb{R}^n)$, by showing that it is continuous and tends to 0 at infinity. The inequality $\|S_t f\| \leq \|f\|$ is obvious from the definition of S_t.

Continuity: Choose $x, y \in \mathbb{R}^n$ and consider the processes X_t^x and X_t^y. The basic estimate in Section 6.1 yields

$$\mathbb{E}\|X_t^x - X_t^y\|^2 \leq 2\left(\|x-y\|^2 + A\int_0^t \mathbb{E}\|X_s^x - X_s^y\|^2 ds\right)$$

for some constant A. Putting $F(t) := \int_0^t \mathbb{E}\|X_s^x - X_s^y\|^2 ds$, we may write this as

$$\frac{d}{dt}F(t) \leq 2\|x-y\|^2 + 2AF(t).$$

So

$$\frac{d}{dt}\left(e^{-2At}F(t)\right) \leq 2\|x-y\|^2 e^{-2At}.$$

Hence, as $F(0) = 0$,

$$e^{-2At}F(t) \leq \|x-y\|^2 \frac{1}{A}\left(1 - e^{-2At}\right),$$

or

$$AF(t) \leq \|x-y\|^2\left(e^{2At} - 1\right).$$

We thus obtain that

$$\mathbb{E}\|X_t^x - X_t^y\|^2 \leq 2\|x-y\|^2 + 2AF(t) \leq 2\|x-y\|^2 e^{2At}. \tag{7.6}$$

Now choose $\varepsilon > 0$. Since f is uniformly continuous on $\mathbb{R}^n$, there exists a $\delta' > 0$ such that

$$\forall x', y' \in \mathbb{R}^n : \qquad \|x' - y'\| \leq \delta' \implies |f(x') - f(y')| \leq \frac{\varepsilon}{2}.$$

It follows that for $x, y \in \mathbb{R}^n$ not more than $\delta := \delta'\sqrt{\varepsilon}e^{-At}/2\sqrt{2\|f\|}$ apart we may estimate

$$\begin{aligned}
\Big|(S_t f)(x) - (S_t f)(y)\Big| &= \Big|\mathbb{E}(f(X_t^x)) - \mathbb{E}(f(X_t^y))\Big| \\
&\leq \mathbb{E}\,|f(X_t^x) - f(X_t^y)| \\
&\leq \frac{\varepsilon}{2} + 2\|f\|\ \mathbb{P}\left[\|X_t^x - X_t^y\| > \delta'\right] \\
&\leq \frac{\varepsilon}{2} + 4\|f\|\ \frac{1}{(\delta')^2}e^{2At}\|x-y\|^2 \\
&< \varepsilon,
\end{aligned}$$

where the third inequality uses the Markov inequality and (7.6). We conclude that $S_t f$ is uniformly continuous as well.

Approach to 0 at infinity: It suffices to prove that there exists a $\tau > 0$ such that for all $t \in [0, \tau]$ we have

$$\lim_{\|x\|\to\infty} (S_t f)(x) = 0.$$

Indeed, the same then holds for all t by the semigroup property. Heuristically speaking, we must prove that the diffusion cannot escape to infinity in a finite time. We start with the estimate

$$\mathbb{E}\left\|X_t^{x,(k+1)} - X_t^{x,(k)}\right\|^2 \leq \frac{A^k t^k}{k!} Dt(1 + \|x\|)^2$$

for the iterates of the map $X \mapsto \tilde{X}$ in Section 6.1, starting from the constant process $X_t^{x,0} = x$. Since $X_t^x = x + \sum_k (X_t^{x,(k+1)} - X^{x,(k)})$ (because $(X_t^{x,(k)} \to X_t^x$ as $k \to \infty$), it follows that

$$\mathbb{E}\,\|X_t^x - x\|^2 \leq D_t(1 + \|x\|)^2 \tag{7.7}$$

with $D_t = Dt(\sum_k \sqrt{A^k t^k/k!})^2$. Now choose $\varepsilon > 0$. Let $M > 0$ be such that $|f(x')| \leq \frac{\varepsilon}{2}$ for all $x' \in \mathbb{R}^n$ with $\|x'\| \geq M$, and let $\tau > 0$ be such that $16D_\tau\|f\| < \frac{1}{2}\varepsilon$. Then, by the triangle inequality in $\mathbb{R}^n$ and by Chebyshev's inequality, we have for all $x \in \mathbb{R}^n$ with $\|x\| > 2M \vee 1$ and all $t \in [0, \tau]$:

$$\begin{aligned}
\|(S_t f)(x)\| &\leq \frac{\varepsilon}{2} + \|f\|\ \mathbb{P}\left[\|X_t^x\| < M\right] \\
&\leq \frac{\varepsilon}{2} + \|f\|\ \mathbb{P}\left[\|X_t^x - x\| > \tfrac{1}{2}\|x\|\right] \\
&\leq \frac{\varepsilon}{2} + \|f\|\ \frac{4}{\|x\|^2}\mathbb{E}\|X_t^x - x\|^2 \\
&\leq \frac{\varepsilon}{2} + \|f\|\ \frac{4}{\|x\|^2}D_\tau(1 + \|x\|)^2 \\
&\leq \frac{\varepsilon}{2} + \|f\|\ 4D_\tau\left(\frac{1}{\|x\|} + 1\right)^2 \\
&< \varepsilon.
\end{aligned}$$

Since ε is arbitrary, this proves the claim.

2. From the initial condition $X_0^x = x$ it is obvious that $(S_0 f)(x) = \mathbb{E}(f(X_0^x)) = f(x)$.

3. The semigroup property of S_t is a consequence of the Markov property of X_t. Namely, for all $s, t \geq 0$ and all $x \in \mathbb{R}^n$:

$$\begin{aligned}(S_{t+s} f)(x) &= \mathbb{E}\left(f\left(X_{t+s}^x\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(f\left(X_{t+s}^x\right)|\mathcal{F}_t\right)\right) \\ &= \mathbb{E}\left(\left(S_s f\right)\left(X_t^x\right)\right) \\ &= \left(S_t\left(S_s f\right)\right)(x).\end{aligned}$$

4. To prove the joint continuity of the map $(t, f) \mapsto S_t f$, it in fact suffices to show that $S_t f \longrightarrow f$ as $t \downarrow 0$. Indeed, for all $s > t \geq 0$,

$$\|S_t f - S_s g\| = \|S_t(f - S_{s-t} g)\| \leq \|f - S_{s-t} g\| \leq \|f - g\| + \|g - S_{s-t} g\|.$$

Note that from (7.7) it follows that $\lim_{t\downarrow 0} \mathbb{E}\|X_t^x - x\|^2 = 0$. Now again take $\varepsilon > 0$. Let $\delta > 0$ be such that $\|x - y\| < \delta \Longrightarrow |f(x) - f(y)| < \varepsilon/2$. Let $t_0 > 0$ be such that $\mathbb{E}\|X_t^x - x\|^2 < \varepsilon\delta^2/4\|f\|$ for $0 \leq t \leq t_0$. Then for $t \in [0, t_0]$ we have, again by Chebyshev's inequality,

$$\begin{aligned}|(S_t f)(x) - f(x)| &= \left|\int_\Omega \left(f\left(X_t^x(\omega)\right) - f(x)\right) \mathbb{P}(d\omega)\right| \\ &\leq \frac{\varepsilon}{2} + 2\|f\|\,\mathbb{P}\left[\|X_t^x(\omega) - x\| \geq \delta\right] \\ &\leq \frac{\varepsilon}{2} + \frac{2\|f\|}{\delta^2}\mathbb{E}\|X_t^x - x\|^2 \\ &< \varepsilon.\end{aligned}$$

□

7.3 Generalities on generators

We next describe some basic theory for one-parameter semigroups on Banach spaces. (We refer to the book of E.B. Davies (1980) for further details.)

Definition. The domain of a generator A of a one-parameter semigroup $(S_t)_{t\geq 0}$ on a Banach space $\mathcal{B}$ is defined by

$$\mathrm{Dom}(A) := \left\{ f \in \mathcal{B} \,\middle|\, \lim_{t\downarrow 0} \tfrac{1}{t}\left(S_t f - f\right) \text{ exists} \right\},$$

and for f in this domain, Af denotes the limit.

This definition leads to the following.

Proposition 7.4 *Let A be the generator of a jointly continuous one-parameter semigroup $(S_t)_{t\geq 0}$ on a Banach space $\mathcal{B}$. Then*

(i) $\mathrm{Dom}(A)$ *is dense in* $\mathcal{B}$*;*

(ii) S_t *leaves* $\mathrm{Dom}(A)$ *invariant:* $S_t(\mathrm{Dom}(A)) \subset \mathrm{Dom}(A)$*;*

(iii) $\forall f \in \mathrm{Dom}(A)$: $S_t A f = A S_t f = \frac{d}{dt} S_t f$.

Proof. (i). For $f \in \mathcal{B}$ and $\varepsilon > 0$ we define

$$f_\varepsilon := \frac{1}{\varepsilon} \int_0^\varepsilon S_t f \, dt.$$

Then $\lim_{\varepsilon \downarrow 0} f_\varepsilon = f$, so $\{ f_\varepsilon | \varepsilon > 0, f \in \mathcal{B} \}$ is dense in $\mathcal{B}$. But we also have

$$\begin{aligned}
\frac{1}{h} (S_h - \mathbb{I}) f_\varepsilon &= \frac{1}{h} (S_h - \mathbb{I}) \frac{1}{\varepsilon} \int_0^\varepsilon S_t f \, dt \\
&= \frac{1}{h\varepsilon} \left(\int_h^{h+\varepsilon} S_t f \, dt - \int_0^\varepsilon S_t f \, dt \right) \\
&= \frac{1}{h\varepsilon} \left(\int_0^h S_t (S_\varepsilon f) dt - \int_0^h S_t f \, dt \right) \\
&\longrightarrow \frac{1}{\varepsilon} (S_\varepsilon f - f) \qquad \text{as } h \downarrow 0.
\end{aligned}$$

So $f_\varepsilon \in \mathrm{Dom}(A)$, and hence $\mathrm{Dom}(A)$ is dense in $\mathcal{B}$.

(ii),(iii). If $f \in \mathrm{Dom}(A)$, then

$$\frac{1}{h} (S_h - \mathbb{I}) (S_t f) = S_t \left(\frac{1}{h} (S_h - \mathbb{I}) (f) \right) \to S_t A f, \qquad \text{as } h \downarrow 0.$$

So $S_t f \in \mathrm{Dom}(A)$ and $A S_t f = S_t A f$. The identity $A S_t f = \frac{d}{dt} S_t f$ follows from the definition of A. □

Since S_t is the solution of the differential equation $\frac{d}{dt} S_t f = A S_t f$ with initial condition $S_0 = \mathbb{I}$, it is customary to write

$$S_t = e^{tA}.$$

The next theorem gives us an explicit formula for the generator of an Itô-diffusion on a large subset of its domain (namely, all functions that are twice continuously differentiable and have compact support).

Theorem 7.5 *Let $(S_t)_{t \geq 0}$ be the one-parameter semigroup on $C_0(\mathbb{R}^n)$ associated to an Itô-diffusion with coefficients b and σ. Let A be the generator of $(S_t)_{t \geq 0}$. Then $C_c^2(\mathbb{R}^n) \subset \mathrm{Dom}(A)$, and for all $f \in C_c^2(\mathbb{R}^n)$:*

$$(Af)(x) = \sum_i b_i(x) \frac{\partial f}{\partial x_i}(x) + \tfrac{1}{2} \sum_{i,j} (\sigma(x)\sigma(x)^*)_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}(x). \tag{7.8}$$

Proof. Denote the r.h.s. of (7.8) by $A'f$. By Itô's formula we have

$$df(X_t) = \sum_i \frac{\partial f}{\partial x_i} (X_t) \, dX_i(t) + \tfrac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} (X_t) \, dX_i(t) dX_j(t).$$

Since $dX_i(t) = b_i(X_t)\,dt + \sum_j \sigma_{ij}(X_t)dB_j(t)$, we have

$$dX_i dX_j = \left(\sum_k \sigma_{ik} dB_k\right)\left(\sum_l \sigma_{jl} dB_l\right) = \left(\sum_k \sigma_{ik}\sigma_{jk}\right) dt = (\sigma\sigma^*)_{ij}\, dt.$$

Using the condition $X_0^x = x$, we conclude that for any $f \in C_c^2(\mathbb{R}^n)$,

$$f(X_t^x) - f(x) = \int_0^t \sum_{i,j} \frac{\partial f}{\partial x_i}(X_s^x)\,\sigma_{ij}(X_s^x)\,dB_j(s) + \int_0^t (A'f)(X_s^x)\,ds. \tag{7.9}$$

The first part on the r.h.s. is a martingale and therefore

$$(Af)(x) = \lim_{t\downarrow 0} \tfrac{1}{t}\left(\mathbb{E}(f(X_t^x)) - f(x)\right) = \lim_{t\downarrow 0} \mathbb{E}\left(\tfrac{1}{t}\int_0^t (A'f)(X_s^x)\,ds\right). \tag{7.10}$$

Since $A'f$ is a continuous function, we have almost surely

$$\lim_{t\downarrow 0} \frac{1}{t}\int_0^t (A'f)(X_s^x)\,ds \longrightarrow (A'f)(x),$$

and since $A'f$ is bounded we can apply dominated convergence to the r.h.s. of (7.10) and conclude that $(Af)(x) = (A'f)(x)$. □

Theorem 7.5 shows that on $C_c^2(\mathbb{R}^n)$ the generator A acts as a second-order differential operator, with a first-order part coming from the drift and a second-order part coming from the diffusion. This fundamental fact forms a bridge between the theory of partial differential equations and the study of diffusions.

In general it is difficult to calculate the evolution semigroup S_t, but in a few exceptional cases, such as that of the Ornstein-Uhlenbeck process, the differential equation $\frac{d}{dt}S_t = S_t A$ can be explicitly solved.

Proposition 7.6 *Let η and σ be positive numbers, and let $(S_t)_{t\geq 0}$ be the semigroup of contractions $C_0(\mathbb{R}) \to C_0(\mathbb{R})$ associated to the stochastic differential equation*

$$dX_t = -\eta X_t\,dt + \sigma\,dB_t. \tag{7.11}$$

Then for all $f \in C_0(\mathbb{R})$ an $x \in \mathbb{R}$,

$$(S_t f)(x) = \frac{1}{\sigma}\sqrt{\frac{\eta}{\pi(1-e^{-2\eta t})}} \int_{-\infty}^{\infty} \exp\left(-\frac{\eta(xe^{-\eta t} - y)^2}{\sigma^2(1-e^{-2\eta t})}\right) f(y)\,dy.$$

Exercise Prove this proposition by checking explicitly that for all $t \geq 0$

$$\frac{d}{dt}S_t f = S_t(Af),$$

where A is the generator of the Ornstein-Uhlenbeck process, given for $f \in C_0^2(\mathbb{R})$ by

$$Af(x) := -\eta x f'(x) + \tfrac{1}{2}\sigma^2 f''(x).$$

Below we shall give a stochastic proof.

Proof. First note that the solution X_t^x of (7.11) with initial condition $X_0^x = x$ is given by

$$X_t^x = xe^{-\eta t} + \sigma N_{f_t} \ ,$$

where $f_t(x) = 1_{[0,t]}(x)e^{\eta(x-t)}$. (Cf. (3.3).) For $t \geq 0$, let γ_t denote the Gaussian density with mean 0 and variance $\sigma^2 \|f_t\|^2 = \sigma^2(1-e^{-2\eta t})/2\eta$, and for $f \in C_0^2(\mathbb{R})$, let $\hat{f}$ denote the Fourier transform of f, so that

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x} \hat{f}(\omega)\, d\omega \ ,$$

Then, since $\mathbb{E}(e^{i\omega\sigma N_{f_t}}) = \hat{\gamma}_t(\omega)$, we have, using Fubini's theorem and the property that the Fourier transform of a convolution product equals the product of the Fourier transforms,

$$\begin{aligned}
(S_t f)(x) &= \mathbb{E}\big(f(X_t^x)\big) \\
&= \frac{1}{2\pi}\mathbb{E}\left(\int_{-\infty}^{\infty} e^{i\omega X_t^x} \hat{f}(\omega) d\omega\right) \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbb{E}\left(e^{i\omega x e^{-\eta t} + i\omega\sigma N_{f_t}}\right) \hat{f}(\omega)\, d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega x e^{-\eta t}} \hat{\gamma}_t(\omega)\hat{f}(\omega)\, d\omega \\
&= \int_{-\infty}^{\infty} \gamma_t(xe^{-\eta t} - y) f(y)\, dy \ .
\end{aligned} \tag{7.12}$$

By writing out the function γ_t we obtain the statement. □

7.4 Dynkin's formula and applications

Having thus identified the generator associated with Itô-diffusions, we next formulate an important formula for *stopped* Itô-diffusions. A *stopping time* τ for $(X_t^x)_{t\geq 0}$ is a random variable such that for all $t \geq 0$ the event $[\tau \leq t] := \{\omega \in \Omega | \tau(\omega) \leq t\}$ is an element of the sigma-algebra $\mathcal{F}_t$ generated by $(X_s^x)_{s\in[0,t]}$.

Theorem 7.7 (Dynkin's formula) *Let X be a diffusion in $\mathbb{R}^n$ with generator A. Let τ be a stopping time with $\mathbb{E}(\tau) < \infty$. Then for all $f \in C_c^2(\mathbb{R}^n)$ and all $x \in \mathbb{R}^n$,*

$$\mathbb{E}^x\left(f\left(X_\tau\right)\right) = f\left(x\right) + \mathbb{E}^x\left(\int_0^\tau \left(Af\right)\left(X_s\right) ds\right) .$$

Proof. The theorem is intuitive, but its proof is rather technical. We limit ourselves here to a very brief sketch.

By applying (7.9) to the stopped process $t \mapsto X_{t\wedge\tau}^x$ and letting t tend to infinity afterwards, we find:

$$f\left(X_\tau^x\right) = f\left(x\right) + \int_0^\tau \sum_{i,j} \frac{\partial f}{\partial x_i}(X_s^x)\sigma_{ij}(X_s^x)dB_j(s) + \int_0^\tau \left(Af\right)\left(X_s^x\right) ds.$$

After taking expectations we get the statement, since stopped martingales starting at 0 have expectation 0. (See Dynkin (1963).) □

The simplest example of a diffusion is Brownian motion itself: $X_t = B_t$ ($b \equiv 0$, $\sigma \equiv id$). Its generator is 1/2 times the Laplacian Δ. We investigate two problems.

Example 7.1.¡ Consider Brownian motion $B_t^a := a + B_t$ starting at $a \in \mathbb{R}^n$. Let $R > \|a\|$. What is the average time B_t^a spends inside the ball $D_R = \{ x \in \mathbb{R}^n : \|x\| \leq R \}$?

Solution: Choose $f \in C_c^2(\mathbb{R}^n)$ such that $f(x) = \|x\|^2$ for $\|x\| \leq R$. Let τ_R^a denote the first time the Brownian motion hits the sphere ∂D_R. Then τ_R^a is a stopping time. Put $\tau := \tau_R^a \wedge T$ and apply Dynkin's formula, to obtain

$$\begin{aligned} \mathbb{E}\left(f\left(B_\tau^a\right)\right) &= f(a) + \mathbb{E}\left(\int_0^\tau \tfrac{1}{2}\left(\Delta f\right)\left(B_s^a\right) ds\right) \\ &= \|a\|^2 + n\mathbb{E}(\tau), \end{aligned}$$

where we use that $\frac{1}{2}\Delta f \equiv n$. Obviously, $\mathbb{E}\left(f\left(B_\tau^a\right)\right) \leq R^2$. Therefore

$$\mathbb{E}(\tau) \leq \tfrac{1}{n}\left(R^2 - \|a\|^2\right).$$

As this holds uniformly in T, it follows that $\mathbb{E}(\tau_R^a) \leq \frac{1}{n}\left(R^2 - \|a\|^2\right) < \infty$. Hence $\tau \to \tau_R^a$ as $T \to \infty$ by monotone convergence. But then we must have $f(B_\tau^a) \to \|B_{\tau_R^a}^a\|^2 = R^2$ as $T \to \infty$, and so in fact

$$\mathbb{E}\left(\tau_R^a\right) = \tfrac{1}{n}\left(R^2 - \|a\|^2\right)$$

by dominated convergence.

Example 7.2.¡ Let $b \in \mathbb{R}^n$ be a point outside the ball D_R. What is the probability that the Brownian motion starting in b ever hits ∂D_R?

Solution: We cannot use Dynkin's formula directly, because we do not know whether the Brownian motion will ever hit ∂D_R (i.e., whether $\tau < \infty$). In order to obtain a well-defined stopping time, we need a bigger ball $D_M := \{ x \in \mathbb{R}^n \underline{| \|x\| \leq M} \}$ enclosing the point b. Let $\sigma_{M,R}^b$ be the first exit time from the annulus $A_{M,R} := \overline{D_M \backslash D_R}$ starting from b. Then clearly $\sigma_{M,R}^b = \tau_M^b \wedge \tau_R^b$. Now take $A = \frac{1}{2}\Delta$, the generator of Brownian motion, and suppose that $f_{M,R}\colon A_{M,R} \to \mathbb{R}^n$ satisfies the following requirements:

(i) $\Delta f_{M,R} = 0$, i.e., $f_{M,R}$ is harmonic,
(ii) $f_{M,R}(x) = 1$ for $\|x\| = R$,
(iii) $f_{M,R}(x) = 0$ for $\|x\| = M$.

We then find

$$\mathbb{P}\left[\left\|B_{\sigma_{M,R}^b}^b\right\| = R\right] = \mathbb{E}\left[f_{M,R}\left(B_{\sigma_{M,R}^b}^b\right)\right] = f_{M,R}(b),$$

where the first equality uses (ii) and (iii) and the second equality uses Dynkin's formula in combination with (i). Incidentally, this equation says that $f_{M,R}$ is uniquely determined if it exists.

Next, we let $M \to \infty$ to obtain

$$\mathbb{P}\left[\tau_R^b < \infty\right] = \mathbb{P}\left[\exists M \geq \|b\| : \tau_R^b < \tau_M^b\right],$$

because any path of Brownian motion that hits ∂D_R must be bounded. From the latter we in turn obtain

$$\begin{aligned}\mathbb{P}\left[\tau_R^b < \infty\right] &= \mathbb{P}\left(\cup_{M \geq \|b\|}[\tau_R^b < \tau_M^b]\right) = \lim_{M\to\infty} \mathbb{P}\left[\tau_R^b < \tau_M^b\right] \\ &= \lim_{M\to\infty} \mathbb{P}\left[\left\|B^b_{\sigma^b_{M,R}}\right\| = R\right] = \lim_{M\to\infty} f_{M,R}(b).\end{aligned}$$

Thus, the only thing that remains to be done is to calculate this limit, i.e., we must solve (i)-(iii).

For $n = 2$ we find, after some calculation:

$$f_{M,R}(b) = \frac{\log\|b\| - \log M}{\log R - \log M} \to 1 \quad \text{as } M \to \infty.$$

For $n \neq 2$, on the other hand, we find:

$$f_{M,R}(b) = \frac{\|b\|^{2-n} - M^{2-n}}{R^{2-n} - M^{2-n}} \to \begin{cases} 1 & \text{if } n = 1 \\ (\|b\|/R)^{2-n} & \text{if } n \geq 3. \end{cases}$$

It follows that Brownian motion with $n = 1$ or $n = 2$ is *recurrent*, since it will hit any sphere with probability 1. But for $n \geq 3$ Brownian motion is *transient*.

8 Transformations of diffusions

In this section we treat two formulas from the theory of diffusions, both of which describe ways to transform diffusions and are proved using Itô-calculus.

8.1 The Feynman-Kac formula

Theorem 8.1 (Feynman-Kac formula) *For $x \in \mathbb{R}^n$, let X_t^x be an Itô-diffusion with generator A and initial condition $X_0^x = x$. Let $v\colon \mathbb{R}^n \to [0,\infty)$ be continuous, and let $S_t^v f$ for $f \in C_0(\mathbb{R}^n)$ and $t \geq 0$ be given by*

$$(S_t^v f)(x) = \mathbb{E}\left(f\left(X_t^x\right) \exp\left(-\int_0^t v(X_u^x) du \right)\right).$$

Then $(S_t^v)_{t\geq 0}$ is a jointly continuous semigroup on $C_0(\mathbb{R}^n)$ (in the sense of Definition 7.2) with generator $A - v$.

Proof. It is not difficult to show, by the techniques used in the proof of Theorem 7.3, that $S_t^v f$ lies again in $C_0(\mathbb{R}^n)$, and that the properties 1, 2 and 4 in Definition 7.2 hold (note that v is non-negative). It is illuminating, however, to explicitly prove Property 3.

For $0 \leq s \leq t$, let

$$Z_{s,t}^x := \exp\left(-\int_s^t v\left(X_u^{s,x}\right) du \right).$$

Property 3 in Definition 7.2 is preserved due to the particular form of the process $Z_{0,t}^x$. In fact, let $f \in C_0(\mathbb{R}^n)$. Then

$$\begin{aligned}
\left(S_{t+s}^v f\right)(x) &= \mathbb{E}\left(Z_{0,t+s}^x f\left(X_{t+s}^x\right)\right) \\
&= \mathbb{E}\left(Z_{0,t}^x Z_{t,t+s}^{X_t^x} f\left(X_{t+s}^x\right)\right) \\
&= \mathbb{E}\left(Z_{0,t}^x \mathbb{E}\left(Z_{t,t+s}^{X_t^x} f\left(X_{t+s}^{t,X_t^x}\right) | \mathcal{F}_t\right)\right) \\
&= \mathbb{E}\left(Z_{0,t}^x g\left(X_t^x\right)\right) \\
&= \left(S_t^v g\right)(x),
\end{aligned}$$

where

$$g(y) = \mathbb{E}\Big(Z_{t,t+s}^y f(X_{t+s}^{t,y}) | \mathcal{F}_t \Big) = \mathbb{E}\Big(Z_{0,s}^y f(X_s^y)\Big) = (S_s^v f)(y)$$

by stationarity. So indeed

$$\left(S_{t+s}^v f\right)(x) = \left(S_t^v \circ S_s^v f\right)(x).$$

Let us finally show that $A - v$ is the generator of $(S_t^v)_{t\geq 0}$. To that end we calculate (dropping the upper index x):

$$\begin{aligned}
d\left(Z_{0,t} f\left(X_t\right)\right) &= f\left(X_t\right) dZ_{0,t} + Z_{0,t} df\left(X_t\right) \\
&= \left[-v\left(X_t\right) f\left(X_t\right) + (Af)\left(X_t\right)\right] Z_{0,t} dt + Z_{0,t} \langle \sigma\left(X_t\right)^T \nabla f\left(X_t\right), dB_t \rangle.
\end{aligned} \tag{8.1}$$

Here we use Itô's formula and $dX_t = b(X_t)dt + \sigma(X_t)dB_t$ to compute

$$
\begin{aligned}
df\left(f(X_t)\right) &= \sum_i \frac{\partial f}{\partial x_i}(X_t)dX_i(t) + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}(X_t)dX_i(t)dX_j(t) \\
&= \sum_i \frac{\partial f}{\partial x_i}(X_t)b_i(X_t)dt + \sum_{i,j}\frac{\partial f}{\partial x_i}(X_t)\sigma_{ij}(X_t)dB_j(t) \\
&\quad \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}(X_t)(\sigma\sigma^*)_{ij}(X_t)dt,
\end{aligned}
\tag{8.2}
$$

where the first and third term are precisely $(Af)(X_t)dt$ according to Theorem 7.5. Taking expectations on both sides of (8.1), we get

$$d\mathbb{E}\left(Z_{0,t}f\left(X_t\right)\right) = \mathbb{E}\left(\left[-vf + Af\right]\left(X_t\right)Z_{0,t}dt\right).$$

In short, $d(S_t^v f) = S_t^v\left((A-v)f\right)dt$, which means that $A - v$ is indeed the generator of the semigroup $(S_t^v)_{t\geq 0}$ (recall Proposition 7.4(iii)). □

We can give the semigroup $(S_t^v)_{t\geq 0}$ a clear probabilistic interpretation. The positive number $v(y)$ is viewed as a "hazard rate" at $y \in \mathbb{R}^n$, the probability per unit time for the diffusion process to be "killed". Let us extend the state space $\mathbb{R}^n$ by a single point ∂, the "coffin" state, where the system ends up after being killed. Then it can be shown that there exists a stopping time τ, the "killing time", such that the process Y_t^x given by

$$Y_t^x := \begin{cases} X_t^x & \text{if } t \leq \tau \\ \partial & \text{if } t > \tau \end{cases}$$

satisfies

$$\mathbb{E}\left(f\left(Y_t^x\right)\right) = \mathbb{E}\left(f\left(X_t^x\right)\exp\left(-\int_0^t v\left(X_u^x\right)du\right)\right),$$

provided we define $f(\partial) := 0$. The proof of this requires an explicit construction of the killing time τ, which we shall not give here.

Example 8.1.¡ The Feynman-Kac formula was originally formulated as a non-rigorous "path-integral formula" in quantum mechanics by R. Feynman, and was later reformulated in terms of diffusions by M. Kac. The connection with quantum mechanics can be stated as follows. If X_t is Brownian motion, then the generator of $(S_t^v)_{t\geq 0}$ is $\frac{1}{2}\Delta - v$. This is $(-1)\times$ the Hamilton operator of a particle in a potential v in $\mathbb{R}^n$. According to Schrödinger, the evolution in time of the wave function $\psi \in L^2(\mathbb{R}^n)$ describing such a particle is given by $\psi \mapsto U_t^v\psi$, where U_t^v is a group of unitary operators given by

$$U_t^v = \exp\left(it\left(\tfrac{1}{2}\Delta - v\right)\right) \qquad (t \in \mathbb{R}).$$

This group can be obtained by analytic continuation in t into the complex domain of the semigroup

$$S_t^v = \exp\left(t\left(\tfrac{1}{2}\Delta - v\right)\right) \qquad (t \in [0,\infty)).$$

Example 8.2.¡ Consider the partial differential equation

$$\frac{\partial}{\partial t}u(x,t) = \Delta u(x,t) + \xi(x)u(x,t), \quad x \in \mathbb{R}^d, t \geq 0, \tag{8.3}$$

with initial conditions

$$u(x,0) = 1 \quad \forall x \in \mathbb{R}^d. \tag{8.4}$$

Here, Δ is the Laplacian acting on the spatial variable and $\{\xi(x) : x \in \mathbb{R}^d\}$ is a field of random variables that plays the role of a random medium. Depending on the choice that is made for the probability law of this random field, (8.3) can be used in various areas of chemistry and biology. For instance, $u(x,t)$ describes the flow of heat in a medium with randomly located sources and sinks.

A formal solution of (8.3) and (8.4) can be written down with the help of the Feynman-Kac formula:

$$u(x,t) = \mathbb{E}\left(\exp\left[\int_0^t \xi(B_s^x)\, ds\right]\right).$$

This representation expresses $u(x,t)$ in terms of a Brownian motion starting at x and is the starting point for a detailed analysis of the behaviour of the random field $\{u(x,t) : x \in \mathbb{R}^d\}$ as $t \to \infty$. (See Carmona and Molchanov (1994)).

Finally, if v fails to be nonnegative, then the Feynman-Kac formula may still hold. For instance, it suffices that v be bounded from below. This guarantees that the properties in Definition 7.2 hold.

8.2 The Cameron-Martin-Girsanov formula

Brownian motion starting at $x \in \mathbb{R}^n$ can be represented naturally on the probability space $(\Omega, \mathcal{F}, \mathbb{P}^x)$, where $\Omega = C([0,T] \to \mathbb{R}^n)$ is the space of its paths, $\mathcal{F}$ is the σ-algebra generated by the cylinder sets $\{\omega \in \Omega \,|\, a \leq \omega(t) \leq b\}$ $(a, b \in \mathbb{R},\ t \in [0,T])$, and $\mathbb{P}^x$ is the probability distribution on $C[0,t]$ of $(B_t^x := x + B_t)_{t\in[0,T]}$ constructed in Section 2. We want to compare the probability distribution $\mathbb{P}^x$ with that of another process derived from it, namely, Brownian motion with a drift.

Let X_t^x denote the solution of

$$dX_t^x = b\left(X_t^x\right) dt + dB_t, \qquad X_0^x = x, \tag{8.5}$$

where b is bounded and Lipschitz continuous. The process X_t^x induces a probability measure $\mathbb{P}_b^x$ on Ω.

Theorem 8.2 (Cameron-Martin-Girsanov formula) *For all $x \in \mathbb{R}^n$ the probability measures $\mathbb{P}^x$ and $\mathbb{P}_b^x$ are mutually absolutely continuous with Radon-Nikodym derivative given by*

$$\frac{d\mathbb{P}_b^x}{d\mathbb{P}^x}\left((B_s^x)_{s\in[0,T]}\right) = \exp\left(\int_0^T \langle b\left(B_s^x\right), dB_s^x\rangle - \frac{1}{2}\int_0^T \|b\left(B_s^x\right)\|^2\, ds\right).$$

Proof. Fix $x \in \mathbb{P}^n$. The idea of the proof is to show that X_t^x has the same distribution under $\mathbb{P}^x$ as B_t^x has under $\rho_T \mathbb{P}^x$, where ρ_T denotes the Radon-Nikodym derivative of the theorem. In other words, we shall show that for all $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq T$ and all $f_1, f_2, \ldots f_n \in C_0(\mathbb{R}^n)$:

$$\mathbb{E}^x \left(f_1(X_{t_1}) \times \cdots \times f_n(X_{t_n})\right) = \mathbb{E}^x \left(\rho_T f_1\left(B_{t_1}^x\right) \times \cdots \times f_n\left(B_{t_n}^x\right)\right). \tag{8.6}$$

Indeed, by the definition of $\mathbb{P}_b^x$ the l.h.s. is equal to

$$\int_\Omega f_1(\omega(t_1)) \times \cdots \times f_n(\omega(t_n)) \mathbb{P}_b^x(d\omega)$$

and the r.h.s. is equal to

$$\int_\Omega f_1\left(\omega(t_1)\right) \times \cdots \times f_n\left(\omega(t_n)\right) \rho_T(\omega) \mathbb{P}^x\left(d\omega\right),$$

while the functions $\omega \mapsto f_1\left(\omega(t_1)\right) \times \cdots \times f_n\left(\omega(t_n)\right)$ generate the σ-algebra $\mathcal{F}$ when $t_1, \ldots, t_n$ are running. We shall prove (8.6) by showing that both sides are equal to

$$S_{t_1}\left(f_1\ S_{t_2 - t_1}\left(\cdots S_{t_n - t_{n-1}}(f_n) \cdots\right)\right)(x). \tag{8.7}$$

Let us start with the l.h.s. of (8.6). First we note that, for all $0 \leq s \leq t \leq T$, all F in the algebra $\mathcal{A}_s$ of bounded $\mathcal{F}_s$-measurable functions on Ω and all $f \in C_0(\mathbb{R}^n)$, by the property of conditional expectations and the Markov property we have that

$$\begin{aligned} \mathbb{E}^x(F\, f(X_t)) &= \mathbb{E}^x\left(\mathbb{E}(F\, f(X_t))|\mathcal{F}_s\right) \\ &= \mathbb{E}^x\left(F\, \mathbb{E}(f(X_t))|\mathcal{F}_s\right) \\ &= \mathbb{E}^x\left(F\, (S_{t-s} f)(X_s)\right). \end{aligned}$$

If we apply this result repeatedly to the product $f_1(X_{t_1}) \times \cdots \times f_n(X_{t_n})$, first projecting onto $\mathcal{A}_{t_{n-1}}$, then onto $\mathcal{A}_{t_{n-2}}$, and all the way down to $\mathcal{A}_{t_1}$, then we obtain (8.7).

Next consider the r.h.s. of (8.6), which is more difficult to handle. We shall show that it is also given by (8.7), in three steps:

Step 1. For all $0 \leq t \leq T$ and $F \in \mathcal{A}_t$:

$$\mathbb{E}^x\left(F \rho_T\right) = \mathbb{E}^x\left(F \rho_t\right). \tag{8.8}$$

Hence the r.h.s. of (8.6) does not depend on T (as long as $T \geq t_n$).

Step 2. For all $0 \leq s \leq T$, $F \in \mathcal{A}_s$ and $f \in C_0(\mathbb{R}^n)$:

$$\mathbb{E}^x\left(F\, f\left(B_t\right) \rho_T\right) = \mathbb{E}^x\left(F\, (S_{t-s} f)(B_s) \rho_T\right). \tag{8.9}$$

Step 3. We apply (8.9) repeatedly, first with $t = t_n$, $s = t_{n-1}$, $F \in \mathcal{A}_{t_{n-1}}$, then with $t = t_{n-1}$, $s = t_{n-2}$, $\mathcal{F} \in \mathcal{A}_{t_{n-2}}$, continuing all the way down to $t = t_1$, and we obtain (8.7).

Thus, to complete the proof it remains to prove (8.8) and (8.9).

Proof of 1. Put $\rho_t := \exp(Z_t)$ with

$$dZ_t = \langle b(B_t), dB_t\rangle - \tfrac{1}{2}\|b(B_t)\|^2\, dt.$$

Then ρ_T is as defined above and Itô's formula gives

$$d\rho_t = \exp(Z_t)\, dZ_t + \tfrac{1}{2}\exp(Z_t)(dZ_t)^2 = \rho_t\langle b(B_t), dB_t\rangle,$$

where two terms cancel because $(dZ_t)^2 = \|b(B_t)\|^2$. It follows that $t \mapsto \rho_t$ is a martingale. Therefore, for $F \in \mathcal{A}_t$,

$$\begin{aligned}\mathbb{E}^x(F\rho_T) &= \mathbb{E}^x(\mathbb{E}(F\rho_T|\mathcal{F}_t)) \\ &= \mathbb{E}^x(F\mathbb{E}(\rho_T|\mathcal{F}_t)) \\ &= \mathbb{E}^x(F\rho_t).\end{aligned}$$

Proof of 2: Note that, by Itô's formula,

$$\begin{aligned}d(\rho_t f(B_t)) &= f(B_t)\, d\rho_t + \rho_t df(B_t) + (d\rho_t)(df(B_t)) \\ &= \rho_t\Big(f(B_t)\langle b(B_t), dB_t\rangle + \langle\nabla f(B_t), dB_t\rangle\Big) \\ &\quad + \rho_t\left(\tfrac{1}{2}\Delta f(B_t)\, dt + \langle b(B_t), \nabla f(B_t)\rangle dt\right)\end{aligned}$$

following a calculation similar to (8.1) and (8.2). Hence, for $0 \le s \le t$ (with $F \in \mathcal{A}_s$) we have, by (8.8),

$$\begin{aligned}\frac{d}{dt}\mathbb{E}^x(F\rho_T f(B_t)) &= \frac{d}{dt}\mathbb{E}^x(F\rho_t f(B_t)) \\ &= \mathbb{E}^x(F\rho_t(Af)(B_t)) \\ &= \mathbb{E}^x(F(Af)(B_t)\rho_T),\end{aligned}$$

where $A := \frac{1}{2}\Delta + \langle b, \nabla\rangle$ is the generator of (X_t). Therefore the l.h.s. and the r.h.s. of (8.9) have the same derivative with respect to t for any $t \ge s \ge 0$. Since they are equal for $t = s$, they must be equal for all $t \ge s \ge 0$. □

The Cameron-Martin-Girsanov formula is a generalisation by Girsanov (1960) of the original formula of Cameron and Martin (1944) for a translation in Wiener space, which we now give as a special case. Let h be a square integrable function $[0, T] \to \mathbb{R}$, and consider the shifted noise $\tilde{N}$ given by

$$\tilde{N}_f := \langle h, f\rangle + N_f\ .$$

Clearly this corresponds to a shift in the associated Brownian motion given by

$$\tilde{B}_t = \int_0^t h(s)ds + B_t\ .$$

Now, since this shifted Brownian motion satisfies the stochastic differential equation

$$d\tilde{B}_t = h(t)\, dt + dB_t\ ,$$

the Cameron-Martin-Girsanov formula says that the probability distributions $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are related by

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \exp\left(\int_0^T h(t)\, dB_t - \tfrac{1}{2}\int_0^T h(t)^2\, dt\right) = e^{N_h - \frac{1}{2}||h||^2}\ .$$

This Radon-Nikodym derivative can be understood as a quotient of Gaussian densities, as the following analogy shows. Let γ be the standard Gaussian density on the real line, and γ_μ is its translate over a distance $\mu \in \mathbb{R}$, then

$$\frac{\gamma_\mu(x)}{\gamma(x)} = \frac{e^{-\frac{1}{2}(x-\mu)^2}}{e^{-\frac{1}{2}x^2}} = e^{2\mu x - \frac{1}{2}\mu^2}\ .$$

9 The linear Kalman-Bucy filter

The Kalman-Bucy filter is an algorithm that filters out random errors occurring in the observation of a stochastic process. When put on a fast computer, the algorithm follows the observations in "real time": at each moment it produces the best possible estimate of the actual value of the process under observation (given certain probabilistic assumptions on the process itself and on the observation errors).

The filter is called "linear" because the estimates depend linearly on the observations. In fact, the model on which the algorithm is based is situated in a Hilbert space of jointly Gaussian random variables with mean zero. In such a space independence is equivalent to orthogonality. For the general nonlinear theory of stochastic filtering we refer to the book of Kallianpur (1980).

The basic idea of linear filtering is best understood by a simple example. From there we shall build up the full model of Kalman and Bucy.

9.1 Example 1: observation of a single random variable

Let X be a Gaussian random variable with mean 0 and variance a^2. Suppose that we cannot observe X directly, but only with some small Gaussian error W, independent of X, that has mean 0 and variance m^2. So we observe

$$Z := X + W.$$

We are interested in the best estimate $\hat{X}$ of X based on our observation of Z and the above assumptions. It is reasonable to interpret "best" as the requirement that $\mathbb{E}((X-\hat{X})^2)$ be minimal, which leads to

$$\hat{X} := \mathbb{E}\left(X|\mathcal{F}(Z)\right),$$

i.e., the orthogonal projection of X onto $L^2(\Omega, \mathcal{F}(Z), \mathbb{P})$, where $\mathcal{F}(Z)$ is the σ-algebra generated by Z. This is equivalent to projecting X onto the one-dimensional space $\mathbb{R}Z$.

Proposition 9.1 *The best estimate of X is given by (see Fig. 5)*

$$\hat{X} = \frac{\mathbb{E}(XZ)}{\mathbb{E}(Z^2)}Z = \frac{a^2}{a^2+m^2}Z.$$

Proof. Put

$$Y := \frac{a^2}{a^2+m^2}Z.$$

Then, because $\mathbb{E}(XW) = 0$,

$$\mathbb{E}\big((X-Y)Z\big) = \mathbb{E}\left(\left(\frac{m^2X - a^2W}{a^2+m^2}\right)(X+W)\right) = 0.$$

So $(X-Y) \perp Z$. For jointly Gaussian random variables with mean zero, orthogonality (i.e., absence of correlation) implies independence. So $(X-Y) \perp\!\!\!\perp Z$. It follows that $X - Y \perp\!\!\!\perp F$ for

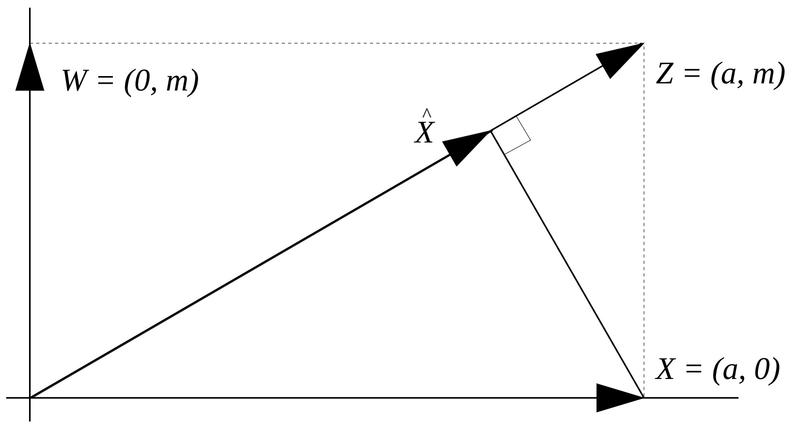


Figure 5: Relative positions of X, W, Z and $\hat{X}$ in the space spanned by X and W. Orthogonality corresponds to independence. Distances in the plane correspond to distances in mean square.

all $F \in L^2(\Omega, \mathcal{F}(Z), \mathbb{P})$, and since $Y \in L^2(\Omega, \mathcal{F}(Z), \mathbb{P})$ it follows that $Y = \mathbb{E}(X|\mathcal{F}(Z)) =: \hat{X}$ (see Fig. 5). □

The mean square distance between X and $\hat{X}$ can be read off from Fig. 5:

$$\|X - \hat{X}\|^2 = \|X\|^2 - \|\hat{X}\|^2 = a^2 - \frac{a^4}{a^2 + m^2} = \frac{a^2 m^2}{a^2 + m^2} = \left(\frac{1}{a^2} + \frac{1}{m^2}\right)^{-1}.$$

Note that $\|X - \hat{X}\| \leq \min(a, m)$.

9.2 Example 2: repeated observation of a single random variable

Next, let us measure X several times with i.i.d. Gaussian observation errors $W_1, W_2, \ldots, W_k$, all independent of X. This gives us the sequence of outcomes

$$Z_j = X + W_j, \qquad (j = 1, \cdots, k).$$

Again we ask for the best estimate $\hat{X}_k$ based on these observations.

Theorem 9.2 *The best estimate of X after k observations is given by*

$$\hat{X}_k := \mathbb{E}(X|\mathcal{F}(Z_1, \ldots, Z_k)) = \frac{a^2}{a^2 + \frac{1}{k}m^2}\left(\frac{1}{k}\sum_{j=1}^{k} Z_j\right).$$

Note that $\hat{X}_k \to X$ almost surely as $k \to \infty$, because $\frac{1}{K}\sum_{j=1}^{k} W_j \longrightarrow 0$ almost surely as $k \longrightarrow \infty$. Thus, frequent observation allows us to retrieve X without error.

We shall prove Theorem 9.2 in two ways. First we shall give a direct proof. After that we shall give a recursive proof suggesting a way to implement this estimation procedure in a machine in "real time".

Direct Proof. As the problem is linear, and as there is no reason to prefer any one of the Z_j to another, let us try as our best estimate

$$Y_k = \alpha_k \sum_{j=1}^{k} Z_j$$

for some constant α_k. Again, we require $\mathbb{E}((X - Y_k)Z_i) = 0$ $(i = 1, \ldots, k)$. So, because $\mathbb{E}(XZ_i) = \mathbb{E}(X^2) = a^2$,

$$0 = \mathbb{E}((X - Y_k)\, Z_i) = \mathbb{E}\left(\left(X - \alpha_k \sum_{j=1}^{k} Z_j\right) Z_i\right) = a^2 - \alpha_k \sum_{j=1}^{k} \mathbb{E}(Z_i Z_j)\,.$$

Now

$$\mathbb{E}(Z_i Z_j) = \mathbb{E}\big((X + W_i)(X + W_j)\big) = \begin{cases} a^2 & \text{if } i \neq j; \\ a^2 + m^2 & \text{if } i = j, \end{cases}$$

which gives

$$\sum_{j=1}^{k} \mathbb{E}(Z_i Z_j) = ka^2 + m^2.$$

Using this result we obtain

$$\alpha_k = \frac{a^2}{ka^2 + m^2} = \frac{1}{k} \frac{a^2}{a^2 + \frac{1}{k} m^2}.$$

Therefore $Y_k = \hat{X}_k$ if α_k is given the above value. □

Recursive Proof. The goal of this proof is to find a recurrence relation in k for $\hat{X}_k$. This is done by means of a procedure that is essential to linear filtering: one identifies what is *new* in each of the consecutive observations by means of the so-called *innovation process* $(N_j)_{j \geq 1}$. In terms of P_j, the projection onto $\mathcal{Z}_j := \text{span}(Z_1, \ldots, Z_j)$ in the spirit of Fig.5:

$$P_j F = \mathbb{E}(F | \mathcal{F}(Z_1, \ldots, Z_j)),$$

N_j is defined as follows:

$$N_j := Z_j - P_{j-1}(Z_j) = Z_j - P_{j-1}(X + W_j) = Z_j - P_{j-1}(X) = Z_j - \hat{X}_{j-1}.$$

The $N_1, \ldots, N_{j-1}$ all lie in the space $\mathcal{Z}_{j-1}$, whereas N_j is orthogonal to it. It follows that the N_j are orthogonal. Since $\dim(\mathcal{Z}_j) = j$, the variables $N_1, \ldots, N_j$ form an orthogonal basis of $\mathcal{Z}_j$. Therefore

$$\hat{X}_k := P_k X = \sum_{j=1}^{k} \frac{\mathbb{E}(X N_j)}{\mathbb{E}(N_j^2)} N_j \tag{9.1}$$

with the coefficients determined by the orthogonality property. Now,

$$\begin{aligned}
\mathbb{E}(XN_j) &= \mathbb{E}\Big(X\left(X + W_j - \hat{X}_{j-1}\right)\Big) \\
&= \mathbb{E}\Big(X\left(X - \hat{X}_{j-1}\right)\Big) \\
&= \mathbb{E}\left(\left(X - \hat{X}_{j-1}\right)^2\right) =: \sigma_{j-1}^2
\end{aligned}$$

($\hat{X}_{j-1} \perp X - \hat{X}_{j-1}$ because $\hat{X}_{j-1} = P_{j-1}X$) and

$$\mathbb{E}\left(N_j^2\right) = \mathbb{E}\left(\left(X + W_j - \hat{X}_{j-1}\right)^2\right) = m^2 + \sigma_{j-1}^2,$$

(because $W_j \perp X$ and $W_j \perp \hat{X}_{j-1}$). This gives us a recursion for $\hat{X}_k$:

$$\begin{aligned}
\hat{X}_{k+1} &= \hat{X}_k + \frac{\sigma_k^2}{m^2 + \sigma_k^2}\left(Z_{k+1} - \hat{X}_k\right) \\
&= \left(\frac{m^2}{m^2 + \sigma_k^2}\right)\hat{X}_k + \left(\frac{\sigma_k^2}{m^2 + \sigma_k^2}\right) Z_{k+1},
\end{aligned}$$

where the r.h.s. is a convex combination. We are able to calculate σ_k^2 by means of the recursion

$$\begin{aligned}
\sigma_k^2 - \sigma_{k+1}^2 &= \|X - \hat{X}_k\|^2 - \|X - \hat{X}_{k+1}\|^2 = \|X - P_k X\|^2 - \|X - P_{k+1}X\|^2 \\
&= \|P_k X - P_{k+1}X\|^2 = \|\hat{X}_{k+1} - \hat{X}_k\|^2 = \frac{\mathbb{E}\left(XN_{k+1}\right)^2}{\mathbb{E}(N_{k+1}^2)} = \frac{\sigma_k^4}{m^2 + \sigma_k^2}.
\end{aligned}$$

This equation can be simplified by changing to the reciprocal of σ_{k+1}:

$$\frac{1}{\sigma_{k+1}^2} = \frac{1}{\sigma_k^2}\left(1 + \frac{\sigma_k^2}{m^2}\right) = \frac{1}{\sigma_k^2} + \frac{1}{m^2},$$

where $\sigma_0^2 = a^2$. We thus find for σ_k the relation

$$\sigma_k^2 = \frac{m^2 a^2}{m^2 + ka^2}. \tag{9.2}$$

If (9.2) is substituted into (9.2), then we find the following recursion relation for $\hat{X}_{k+1}$:

$$\hat{X}_{k+1} = \left(\frac{m^2 + ka^2}{m^2 + (k+1)\,a^2}\right)\hat{X}_k + \left(\frac{a^2}{m^2 + (k+1)\,a^2}\right) Z_{k+1}. \tag{9.3}$$

This relation expresses the new estimate as a convex combination of the old estimate and the new observation, which is clearly useful for calculation in real time. □

9.3 Example 3: continuous observation of an Ornstein-Uhlenbeck process

We now apply the above recursive program to the observation Z_t of an Ornstein-Uhlenbeck process X_t:

$$dX_t = -\alpha X_t dt + \beta dB_t \quad (X_0 \text{ Gaussian}, X_0 \perp\!\!\!\perp B_t) \tag{9.4}$$

and

$$dZ_t = \gamma X_t dt + \delta dW_t \quad (Z_0 = 0). \tag{9.5}$$

Here B_t and W_t are independent Brownian motions, and $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ are parameters. Note that W_t plays the same role as the i.i.d. random variables W_j in the previous example, and Z_t plays the role of Z_j. The parameters α and β are the drift and the diffusion coefficients of the observed process X_t. We can think of γ as some kind of coupling constant that indicates how strong the process Z_t is influenced by X_t. The parameter δ plays the role of m.

Theorem 9.3 (Linear Kalman-Bucy filter) *Let the Ornstein-Uhlenbeck process X_t and its observation process Z_t be given by (9.4) and (9.5). Then the projection $\hat{X}_t$ of X_t onto $\mathcal{Z}_t$, the closed linear span of $(Z_s)_{0 \le s \le t}$, satisfies the stochastic differential equation*

$$d\hat{X}_t = \left(-\alpha - \frac{\gamma^2}{\delta^2} \sigma(t)^2 \right) \hat{X}_t dt + \frac{\gamma}{\delta^2} \sigma(t)^2 \, dZ_t. \tag{9.6}$$

where $\sigma(t)^2$ satisfies the ordinary differential equation

$$\frac{d}{dt} \sigma(t)^2 = -2\alpha \sigma(t)^2 + \beta^2 - \frac{\gamma^2}{\delta^2} \sigma(t)^4 . \tag{9.7}$$

Equation (9.6) is the analogue of the recursion (9.3) in Example 2 above. Equation (9.7) will be solved at the end of this section. We shall build up the proof of Theorem 9.3 via several lemmas.

In order to calculate $\hat{X}_t$, we first identify the innovation process of Z_t as the closest analogue to an orthogonal basis. This goes as follows. Let

$$V_t := W_t + \frac{\gamma}{\delta} \int_0^t \left(X_s - \hat{X}_s \right) ds. \tag{9.8}$$

Then, since $Z_t = \delta W_t + \gamma \int_0^t X_s ds$, we can write

$$V_t = \frac{1}{\delta} \left(Z_t - \gamma \int_0^t \hat{X}_s ds \right).$$

Hence

$$\delta dV_t = dZ_t - \gamma \hat{X}_t dt = (\mathbb{1} - P_t) \, dZ_t, \tag{9.9}$$

provided we take P_t (the projection onto $\mathcal{Z}_t$) "just before" dt, so that $P_t dW_t = 0$. Because of (9.9), V_t indeed deserves the name of innovation process: it isolates from the observations that part which is independent of the past.

The idea of the proof of Theorem 9.3 is now to write $\hat{X}_t$ as a stochastic integral over the innovation process V_t, analogous to the expansion (9.1) in the orthogonal basis (N_j). This will be done in Lemma 9.5 and requires the following fact:

Lemma 9.4 *V_t is a Brownian motion.*

Proof. The family $\{V_t\}_{t\in[0,T]}$ is jointly Gaussian. It therefore suffices to show that the covariance is the same as for a Brownian motion (cf. (2.4)): $\mathbb{E}(V_tV_s) = s \wedge t$ $(s,t \geq 0)$.

First note that $V_t \in \mathcal{Z}_t$, and that $V_s - V_t \perp \mathcal{Z}_t$ for $s \geq t$ because $V_s - V_t = \int_t^s dV_u = \int_t^s [dW_u + \frac{\gamma}{\delta}(X_u - \hat{X}_u)du]$ is orthogonal to $\mathcal{Z}_t$ by construction. Taking expectations we get

$$\mathbb{E}(V_tV_s) - \mathbb{E}(V_t^2) = \mathbb{E}(V_t(V_s - V_t)) = 0,$$

So $\mathbb{E}(V_tV_s) = \mathbb{E}(V_t^2)$. We next calculate the derivative of V_t^2:

$$d(V_t^2) = 2V_t dV_t + (dV_t)^2 = 2V_t dV_t + dt.$$

Taking expectations we get

$$d\mathbb{E}\left(V_t^2\right) = 2\mathbb{E}(V_t dV_t) + dt = dt,$$

because $V_t \perp\!\!\!\perp dV_t$. Integration now yields the desired result:

$$\mathbb{E}(V_tV_s) = \mathbb{E}(V_t^2) = t = t \wedge s \qquad (s \geq t).$$

□

We now make the following claim, which is the analogue of (9.1).

Lemma 9.5 *For all $F \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, the projection P_t onto $\mathcal{Z}_t$ acting on F is given by*

$$P_tF = \int_0^t f'(s)dV_s,$$

where $f(s) := \mathbb{E}(FV_s)$. In particular, $f' \in L^2([0,t])$.

The proof is deferred to later. An explanation for this theorem can be given by comparing the integral with the sum

$$P_kF = \sum_{j=1}^k \mathbb{E}(FN_j)N_j,$$

where (N_j) is an orthonormal basis ($\mathbb{E}(N_j^2) = 1$). The continuous analogue is the formal expression

$$\begin{aligned} P_tF &= \int_0^t \mathbb{E}(FN_s)N_s ds \qquad (\text{with } N_s = \tfrac{dV_s}{ds}) \\ &= \int_0^t \left(\frac{d}{ds}\mathbb{E}(FV_s)\right) dV_s. \end{aligned}$$

To prove Lemma 9.5 we need the following technical lemma:

Lemma 9.6 *The following three spaces are equal for all $t \geq 0$:*

(i) $\mathcal{Z}_t$,
(ii) $\{ \int_0^t f(s)\, dZ_s | f \in L^2([0,t]) \}$,
(iii) $\{ \int_0^t g(s)\, dV_s | g \in L^2([0,t]) \}$.

Proof. 1. First we show that (ii) is a closed linear space. Let Z be the operator $L^2([0,t]) \to L^2(\Omega, \mathbb{P})$ defined as $Zf = \int_0^t f(s) dZ_s$. Then, because $(dZ_s)^2 = \delta^2 ds$,

$$\delta^2 \|f\|^2 \le \|Zf\|^2 \le \left(\delta^2 + \gamma^2 \mathbb{E}\left(\int_0^t X_s^2 ds\right)\right) \|f\|^2 .$$

Indeed, by the Itô isometry and the independence of X and W, we have

$$\begin{aligned}\mathbb{E}\left(\left(\int_0^t f(s)\, dZ_s\right)^2\right) &= \mathbb{E}\left(\left(\gamma \int_0^t f(s)\, X_s ds + \delta \int_0^t f(s)\, dW_s\right)^2\right) \\ &= \gamma^2 \mathbb{E}\left(\int_0^t f(s)\, X_s ds\right)^2 + \delta^2 \|f\|^2 .\end{aligned}$$

The first term in the r.h.s. lies between 0 and

$$\gamma^2 \left(\int_0^t f(s)^2\, ds\right) \mathbb{E}\left(\left(\int_0^t X_s^2 ds\right)\right).$$

This implies that $Z\colon L^2([0,t]) \to \mathrm{Ran}(Z)$ is bounded and has a bounded inverse, so $\mathrm{Ran}(Z)$ is closed.

2. Clearly, $\mathcal{Z}_t$ is contained in (ii) because, for all $0 \le s \le t$,

$$Z_s = \int_0^t 1_{[0,s]}(u)\, dZ_u = Z\left(1_{[0,s]}\right) \in \mathrm{Ran}(Z).$$

Conversely, suppose that $f = \lim_{j\to\infty} f_j$ with f_j simple. Then $Zf_j \in \mathrm{span}\left((Z_s)_{0\le s\le t}\right)$, which means that Zf_j is a linear combination of Z_s, $0 \le s \le t$. Hence $Zf \in \mathcal{Z}_t$, so that also $\mathcal{Z}_t \supset \mathrm{Ran}(Z)$. Thus, (i)=(ii).

3. Obviously, (iii) is closed because $V\colon L^2([0,t]) \to L^2(\Omega, \mathbb{P}) : g \mapsto \int_0^t g(s)\, dV_s$ is an isometry. We know that $\mathrm{Ran}(V) \subset \mathrm{Ran}(Z)$, since $V_s \in \mathcal{Z}_t = \mathrm{Ran}(Z)$ for all $0 \le s \le t$. To show equality of (i) and (iii) we must show that $Z_s \in \mathrm{Ran}(V)$ for all $0 \le s \le t$. Now, omitting γ and δ for the moment, we may write

$$dZ_s = dV_s + \hat{X}_s ds = dV_s + \left(\int_0^s h_s(r)\, dZ_r\right) ds,$$

for some $h_s \in L^2([0,s])$ because $\hat{X}_s \in \mathrm{span}((Z_r)_{0\le r\le s})$, so that, for all $g \in L^2([0,t])$,

$$\begin{aligned}\int_0^t g(s)\, dV_s &= \int_0^t g(s) dZ_s - \int_0^t \left(\int_0^s h_s(r) dZ_r\right) g(s) ds \\ &= \int_0^t g(r) dZ_r - \int_0^t \left(\int_r^t h_s(r) g(s) ds\right) dZ_r.\end{aligned}$$

By the theory of Volterra integral equations, it is possible for any $u \geq 0$ and $(h_s)_{0 \leq s \leq t}$ to choose g such that

$$g(r) - \int_r^t h_s(r)\, g(s)\, ds = \mathbb{1}_{[0,u]}(r) \qquad (0 \leq r \leq t).$$

Then $Z_u = \int_0^t \mathbb{1}_{[0,u]}(r)\, dZ_r = \int_0^t g(s)\, dV_s \in \mathrm{Ran}(V)$. □

Proof of Lemma 9.5: Use the equality of (i) and (iii) from Lemma 9.6 to obtain a function $g \in L^2([0,t])$ such that

$$P_t F = \int_0^t g(s)\, dV_s.$$

Then, for $0 \leq u \leq t$, we find by the isometric property of integration w.r.t. V_t (Lemma 9.4),

$$\begin{aligned} f(u) &:= \mathbb{E}(F V_u) = \mathbb{E}((P_t F)\, V_u) \\ &= \mathbb{E}\left(\left(\int_0^t g(s)\, dV_s\right)\left(\int_0^t \mathbb{1}_{[0,u]}(r)\, dV_r\right)\right) \\ &= \langle g, \mathbb{1}_{[0,u]} \rangle = \int_0^u g(r)\, dr, \end{aligned}$$

which implies $f' = g$. We thus conclude that $f' \in L^2[0,t]$ and $P_t F = \int_0^t f'(s)\, dV_s$. □

Now we can prove Theorem 9.3.

Proof. By Lemma 9.5 we may write

$$\hat{X}_t = P_t X_t = \int_0^t f_t'(s)\, dV_s,$$

where $f_t(s) := \mathbb{E}(X_t V_s)$. Let us calculate $f_t(s)$ for $0 \leq s \leq t$ using the solution of (9.4) for X_t, which can be written as

$$X_t = \beta \int_0^t \mathrm{e}^{-\alpha(t-s)} dB_s + \mathrm{e}^{-\alpha t} X_0.$$

We have:

$$\begin{aligned} f_t(s) &= \mathbb{E}(X_t V_s) \\ &= \mathbb{E}\left(X_t \left(W_s + \frac{\gamma}{\delta}\int_0^s \left(X_u - \hat{X}_u\right) du\right)\right) \\ &= \frac{\gamma}{\delta}\int_0^s \mathbb{E}\left(X_t \left(X_u - \hat{X}_u\right)\right) du \\ &= \frac{\gamma}{\delta}\int_0^s \mathbb{E}\left(\mathrm{e}^{-\alpha(t-u)} X_u \left(X_u - \hat{X}_u\right)\right) du \\ &= \frac{\gamma}{\delta}\int_0^s \mathrm{e}^{-\alpha(t-u)}\mathbb{E}\left(\left(X_u - \hat{X}_u\right)^2\right) du. \end{aligned}$$

Differentiating this expression w.r.s. s we get

$$f_t'(s) = \frac{\gamma}{\delta} e^{-\alpha(t-s)} \mathbb{E}\left(\left(X_s - \hat{X}_s\right)^2\right) =: \frac{\gamma}{\delta} e^{-\alpha(t-s)} \sigma(s)^2 .$$

This means that $\hat{X}_t$ satisfies the stochastic differential equation:

$$\begin{aligned} d\hat{X}_t &= \left\{ \int_0^t \left[\frac{d}{dt} f_t'(s)\right] dV_s \right\} dt + f_t'(t) dV_t \\ &= -\alpha \hat{X}_t dt + f_t'(t) \left(\frac{1}{\delta} dZ_t - \frac{\gamma}{\delta} \hat{X}_t dt\right) \\ &= \left(-\alpha - \frac{\gamma^2}{\delta^2} \sigma(t)^2\right) \hat{X}_t dt + \frac{\gamma}{\delta^2} \sigma(t)^2 dZ_t, \end{aligned}$$

which is (9.6). To finish the proof of Theorem 9.3 we must find an expression for $\sigma(t)$, given by

$$\sigma(t)^2 := \mathbb{E}\left(\left(X_t - \hat{X}_t\right)^2\right) = \mathbb{E}\left(X_t^2\right) - \mathbb{E}\left(\hat{X}_t^2\right),$$

where (use Lemma 9.5)

$$\mathbb{E}\left(X_t^2\right) = e^{-2\alpha t} \mathbb{E}\left(X_0^2\right) + \beta^2 \int_0^t e^{-2\alpha(t-s)} ds \quad \Longrightarrow \quad \frac{d}{dt} \mathbb{E}\left(X_t^2\right) = -2\alpha \mathbb{E}\left(X_t^2\right) + \beta^2,$$

and

$$\begin{aligned} \frac{d}{dt} \mathbb{E}\left(\hat{X}_t^2\right) &= \frac{d}{dt} \int_0^t f_t'(s)^2 ds \\ &= \int_0^t 2 f_t'(s) \frac{d}{dt} f_t'(s) ds + f_t'(t)^2 \\ &= -2\alpha \int_0^t f_t'(s)^2 ds + \frac{\gamma^2}{\delta^2} \sigma(t)^4 \\ &= -2\alpha \mathbb{E}(\hat{X}_t^2) + \frac{\gamma^2}{\delta^2} \sigma(t)^4 . \end{aligned}$$

It follows that $\sigma(t)^2$ satisfies (9.7), and Theorem 9.3 is proved. □

In the case treated above, where α, β, γ, and δ are constants, (9.7) can be solved (with $\sigma(0) = a$) to give

$$\sigma(t)^2 = \frac{\alpha_1 - K\alpha_2 \exp\left(\gamma^2 (\alpha_2 - \alpha_1) t/\delta^2\right)}{1 - K \exp\left(\gamma^2 (\alpha_2 - \alpha_1) t/\delta^2\right)},$$

where

$$\begin{aligned} \alpha_1 &= \frac{1}{\gamma^2} \left(-\alpha\delta^2 - \delta\sqrt{\alpha^2\delta^2 + \gamma^2\beta^2}\right) \\ \alpha_2 &= \gamma^2 \left(-\alpha\delta^2 + \delta\sqrt{\alpha^2\delta^2 + \gamma^2\beta^2}\right) \end{aligned}$$

and

$$K = \frac{a^2 - \alpha_1}{a^2 - \alpha_2}.$$

This is easily checked by substitution. The formulas are not easy, but explicit.

We note that Theorem 9.3 remains valid when α, β, γ and δ are allowed to depend on time.

10 The Black and Scholes option pricing formula.

Myron Scholes and Fisher Black

In 1973 Black and Scholes published a paper (after two rejections) containing a formula for the fair price of a European call option on stocks. This formula now forms the basis of pricing practice on the option market. It is a fine example of applied stochastic analysis, and marks the beginning of an era were banks employ probabilists that are well versed in Itô calculus.

A (European) call option on stocks is the right, but not the obligation, to buy at some future time T a share of stock at price K. Both T and K are fixed ahead. Since at time T the value of the stock may be higher than K, such a right-to-buy in itself has a value. The problem of this section is: "What is a reasonable price for this right?"

10.1 Stocks, bonds and stock options

Let $(B_t)_{t\in[0,T]}$ be standard Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t\in[0,T]}$. We model the price of one share of a certain *stock* on the financial market by an Itô-diffusion $(S_t)_{t\in[0,T]}$ on the interval $(0, \infty)$, described by the stochastic differential equation

$$dS_t = S_t(\mu dt + \sigma dB_t). \tag{10.1}$$

The constant $\mu \in \mathbb{R}$ (usually positive) describes the relative rate of return of the shares. The constant $\sigma > 0$ is called the *volatility* and measures the size of the random fluctations in the stock value. Shares of stock represent some real asset, for example partial ownership of a company. They can be bought or sold at any time t at the current price S_t.

Let us suppose that on the market certain other securities, called *bonds*, are available that yield a riskless return rate $r \geq 0$. This is comparable to the interest on bank accounts. The value β_t of a bond satisfies the differential equation

$$d\beta_t = \beta_t(rdt). \tag{10.2}$$

The question we are addressing is: How much would we be willing to pay at time 0 for the right to buy at time T one share of stock at the price $K > 0$ fixed ahead? Such a right is called a *European stock option*. The time T is called the *expiry time*, the price K is called the *exercise price*. This option pricing problem turns out to be a problem of stochastic control.

Another type of option is the *American stock option*, where the time at which the shares can be bought is not fixed beforehand but only has an upper limit. The pricing problem for American stock options contains, apart from a stochastic control element, also a halting problem. It is therefore more difficult, and we leave it aside.

The solution to the European option pricing problem is given by the Black and Scholes option pricing formula (10.4) appearing at the end of this Section. As this formula does not look wholely transparant at first sight, we shall introduce the result in three steps, raising the level of complexity slowly by introducing the ingredients one by one.

10.2 The martingale case

Let us first suppose that both the stock price and the bond price are martingales: $\mu = r = 0$ in Equations (10.1) and (10.2). For bonds this simply means that their price is constant: they are the stock-market equivalent of bank notes. For stocks it means that their price behaves like the exponential martingale found in Chapter 5:

$$S_t = S_0 \exp\left(\sigma B_t - \tfrac{1}{2}\sigma^2 t\right).$$

In this stationary world our option pricing problem is easy: a fair price q at time 0 of the right to buy at time T a share of stock at the price K is

$$q := \mathbb{E}\big((S_T - K)^+\big), \tag{10.3}$$

where x^+ stands for $\max(0, x)$. Indeed, if the stock value S_T turns out to be larger than the exercise price K, then the holder of the option makes a profit of $S_T - K$ by buying the share of stock at the price K that he is entitled to, and then immediately selling it again at its current value S_T. On the other hand, if $S_T \leq K$, then his option expires as a worthless contract.

Since we know the price process S_t to be an exponential martingale, we can explicitly evaluate the option price q:

$$\begin{aligned}
q &= \mathbb{E}\big((S_T - K)^+\big) \\
&= \mathbb{E}\left(\left(S_0 e^{\sigma B_T - \frac{1}{2}\sigma^2 T} - K\right)^+\right) \\
&= \int_{-\infty}^{\infty} \left(S_0 e^{w - \frac{1}{2}\sigma^2 T} - K\right)^+ \varphi_{\sigma^2 T}(w)\, dw \\
&= \int_{\log \frac{K}{S_0} + \frac{1}{2}\sigma^2 T}^{\infty} \left(S_0 e^{w - \frac{1}{2}\sigma^2 T} - K\right) \varphi_{\sigma^2 T}(w)\, dw \\
&= S_0\, \Phi_{\sigma^2 T}\left(\log \frac{S_0}{K} + \frac{1}{2}\sigma^2 T\right) - K\, \Phi_{\sigma^2 T}\left(\log \frac{S_0}{K} - \frac{1}{2}\sigma^2 T\right),
\end{aligned} \tag{10.4}$$

where $\Phi_\lambda : u \mapsto \frac{1}{\sqrt{\lambda}}\Phi\left(\frac{u}{\sqrt{\lambda}}\right)$ is the normal distribution fuction with mean 0 and variance λ, and $\varphi_\lambda = \frac{1}{\sqrt{2\pi\lambda}} exp(-w^2/2\lambda)$ is the associated density function. In (10.4) we have made use of the equalities

$$e^{w - \frac{1}{2}\lambda}\varphi_\lambda(w) = \varphi_\lambda(w - \lambda),$$

and

$$\int_x^\infty \varphi_\lambda(w)dw = \Phi_\lambda(-x).$$

The above option pricing formula covers the case $\mu = r = 0$. Surprisingly, it also covers the case $\mu \neq 0, r = 0$, i.e., μ plays no role in the final result. In fact, the full Black and Scholes option pricing formula is obtained by substituting Ke^{-rT} for K to take devaluation into account. It was this surprising disappearance of μ from the formula that caused the difficulties that Black and Scholes experienced in getting their result accepted.

And indeed, for a justification of these statements we need a considerable extension of our theoretical background.

10.3 The effect of stock trading

If we now drop the assumption that $\mu = 0$, but keep $r = 0$, the intuitive argument leading to (10.3) breaks down. Indeed, if $\mu > 0$, then there is an upward drift in the stock value that makes "dollars at time 0" inequivalent to "dollars at time T". Simply building in a discount factor $e^{-\mu t}$ is ill-founded, and will in fact turn out to be incorrect. We shall have to consider seriously the definition of a "fair price". It is reasonable to base such a definition on the question what else could be done with a capital q during the interval of time $[0, T]$. This brings us to the subject of *trading.*

Suppose that a dealer in securities enters the market at time 0 with a capital q. He buys an amount a_0 of stock at the current price S_0, keeping the rest $b_0 := q - a_0 S_0$ in his pocket in the form of banknotes. Then at the times $t_1 < t_2 < t_3 < \cdots < t_n (= T)$ he repeats the following procedure. Let $i \in \{0, 1, 2, \cdots, n-1\}$. At time t_{i+1} our dealer, judging the past stock values S_t, $(0 \leq t \leq t_{i+1})$, decides to change his amount of stock from a_{t_i} to $a_{t_{i+1}}$ by buying or selling. If he chooses to buy more than his capital can pay for, he borrows money, (thus making b_{t+1} negative), and keeps in mind that his loan must be paid back in due course, say after time T. Nevertheless, his tradings must be *self-financing*, i.e., our dealer spends no money, and obtains no money from outside, other than the loans and stock tradings mentioned. So his total capital ("*portfolio value*") just before t_{i+1} must be equal to the portfolio value just after this moment of time:

$$a_{t_i} S_{t_{i+1}} + b_{t_i} = a_{t_{i+1}} S_{t_{i+1}} + b_{t_{i+1}}, \qquad (i = 0, 1, \cdots, n-1). \tag{10.5}$$

Employing the notation of Chapter 2,

$$(\Delta t)_i := t_{i+1} - t_i; \qquad (\Delta X)_i := X_{t_{i+1}} - X_{t_i},$$

we may write (10.5) as

$$(\Delta a)_i (S_{t_i} + (\Delta S)_i) + (\Delta b)_i = 0. \tag{10.6}$$

We note that this equation expresses a kind of "time delay": the amount a_{t_i} of stock bought at time t_i only has its effect at time t_{i+1}. This delay will remain even after the continuous time limit is taken, when it will give rise to an Itô-term.

Now consider the portfolio value C_t at time t:

$$C_t := a_t S_t + b_t.$$

We note that in the case of self-financing trade:

$$\begin{aligned}(\Delta C)_i &:= C_{t_{i+1}} - C_{t_i} \\ &= a_{t_{i+1}} S_{t_{i+1}} + b_{ti+1} - a_{t_i} C_{t_i} - b_{t_i} \\ &= (a_{t_{i+1}} - a_{t_i})(S_{t_{i+1}} - S_{t_i}) + (b_{t_{i+1}} - b_{t_i}) + (a_{t_{i+1}} - a_{t_i})S_{t_i} + a_{t_i}(S_{t_{i+1}} - S_{t_i}) \\ &= (\Delta a)_i(\Delta S)_i + (\Delta b)_i + (\Delta a)_i S_i + a_i(\Delta S)_i\end{aligned}$$

Therefore (10.6) can be written concisely as

$$(\Delta C)_i = a_i(\Delta S)_i.$$

Now, since there is no essential limit to the frequency of trade, the partition of $[0, T]$ generated by the sequence of times $0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = T$ can be made arbitrarily fine. It is therefore reasonable to make the following idealisation.

Definition. A *self-financing trading strategy* is a pair (a, b) of square-integrable adapted processes on $[0, T]$ with continuous paths, such that the sum $C_t := a_t S_t + b_t$ satisfies

$$dC_t = a_t dS_t. \tag{10.7}$$

We are now in a position to define what we mean by the "fair price" of a claim or option. Let $g : [0, \infty) \to [0, \infty)$ be measurable. By a *claim to* $g(S_T)$ we mean the right to cash in at time T the amount $g(S_T)$, which depends on the current stock value S_T at time T.

Definition. A claim to $g(S_T)$ is called *redundant* if there exists a self-financing strategy (a, b) such that with probability 1,

$$C_T := a_T S_T + b_T = g(S_T).$$

By the *fair price* $\mathbb{F}(g(S_T))$ of a redundant claim to $g(S_T)$ we mean

$$\mathbb{F}(g(S_T)) := C_0 = a_0 S_0 + b_0.$$

So $\mathbb{F}(g(S_T)) = q$ if q could be used as the starting capital $q = a_0 S_0 + b_0$ for a trading (a, b) that ends up with $g(S_T)$ with certainty.

10.4 Motivation

In the financial literature the above definition is usually motivated by the following argument (a so-called *arbitrage* argument).

Suppose that claims to $g(S_T)$ were traded at time 0 at a price p higher than q. Then it would be possible to make an unbounded and riskless profit (an "arbitrage") by selling n such claims for the market price p, then to reserve an amount nq as initial capital for the self-financing strategy (na, nb) — yielding with probability 1 the amount $ng(S_T)$ needed to satisfy the claims — and then to pocket the difference $n(p-q)$. Conversely, if the market price p of the claim would be lower than q, then one could buy n claims and apply the strategy $(-na, -nb)$ yielding an immediate gain of $n(q-p)$ at time 0. At time T one could cancel one's debts by executing the n claims to $g(S_T)$. It should be admitted that this second strategy, involving negative share holdings (or *short sales* of stock), is somewhat more artificial than the first. But clearly, the possibility of arbitrage is not fair.

This concludes the motivation of Definition 10.3. In economic theory one often goes one step further and assumes that arbitrage in fact does not occur. It is claimed that the possibility of arbitrage would immediately be used by one of the parties on the market, and this would set the market price equal to the fair price.

10.5 Results

Theorem 10.1 *Let $g \in C_0(0,\infty)$. On a stock market without bonds or interest (i.e., with $r = 0$), the fair price at time 0 of a claim to $g(S_T)$ at time $T > 0$ is*

$$\mathbb{F}(g(S_T)) = \mathbb{E}(g(X_T)),$$

where $(X_t)_{t\in[0,T]}$ is the exponential martingale with parameter σ starting at S_0, i.e., the solution of the stochastic differential equation

$$dX_t = X_t(\sigma dB_t) \qquad \text{with} \quad X_0 = S_0. \tag{10.8}$$

Corollary 10.2 *In the absence of bonds or interest the fair price at time 0 of an option to a share of stock at time T is*

$$\mathbb{F}\left((S_T - K)^+\right) = \mathbb{E}\left((X_T - K)^+\right),$$

where X_t is the exponential martingale of Theorem 10.1. The right hand side is given explicitly by (10.4).

Proof. Approximate the function $x \mapsto (x-K)^+$ by a sequence $g_n \in C_0(0,\infty)$ ad apply Theorem 10.1. □

We now give the proof of Theorem 10.1.

Proof. Let $A := \sigma^2 x^2 \frac{\partial^2}{\partial x}$ denote the generator of the diffusion X_t in (10.8), and define

$$f_t := e^{(T-t)A} g, \qquad (t \in [0,T]).$$

Define a trading strategy (a,b) by

$$a_t := f_t'(S_t) \quad \text{and} \quad b_t := f_t(S_t) - S_t f_t'(S_t).$$

Then $a_T S_T + b_T = f_T(S_T) = g(S_T)$. Moreover, (a,b) is self-financing since the total portfolio value C_t satisfies

$$\begin{aligned} dC_t &:= df_t(S_t) \\ &= \left(\frac{df_t}{dt}\right)(S_t)dt + f_t'(S_t)dS_t + \tfrac{1}{2} f_t''(S_t)(dS_t)^2 \\ &= -(Af_t)(S_t)dt + a_t dS_t + \tfrac{1}{2}\sigma^2 S_t^2 f_t''(S_t)dt \\ &= a_t dS_t. \end{aligned}$$

It follows that the fair price at time 0 is given by

$$\mathbb{F}(g(S_T)) := a_0 S_0 + b_0 = f_0(S_0) = \left(e^{TA} g\right)(S_0) = \mathbb{E}\left(g(X_T^{S_0})\right).$$

□

Note that there is no μ in this proof! Apparently the fair price at time 0 is not influenced by μ To begin to understand this fact, we take the case $g(x) = x$. Clearly the fair price at time 0 of a share at time T is just S_0, not $\mathbb{E}(S_T)$: S_t is automatically a "martingale under $\mathbb{F}$".

Explicit calculation of the strategy for the case $g(x) = (x - K)^+$ of a stock option yields

$$a_t = \Phi_{\sigma^2(T-t)}\left(\log \frac{S_t}{K} + \tfrac{1}{2}\sigma^2(T-t)\right);$$

and

$$b_t = -K\,\Phi_{\sigma^2(T-t)}\left(\log \frac{S_t}{K} - \tfrac{1}{2}\sigma^2(T-t)\right).$$

These expressions describe a smooth steering mechanism, moving from the initial value (a_0, b_0) to the final value (a_T, b_T), given by

$$(a_T, b_T) = \begin{cases} (0,0), & \text{if } S_T \leq K; \\ (1, -K), & \text{if } S_T > K. \end{cases}$$

Thus the pair (a_t, b_t) always moves inside $[0,1] \times [-K, 0]$, the strategy always involves borrowing of money in order to buy up to one single share of stock. In cases where the option is cheap, relatively much has to be borrowed in order to imitate the workings of the option.

10.6 Inclusion of the interest rate

We complete our treatment of the Black and Scholes model by including bonds, securities that yield a riskless return. The presence of these bonds on the market leads to a constant devaluation. Therefore, measured in "dollars at time 0", the exercise price of the option will only measure Ke^{-rT}. Interestingly, this is the only change (in the Black and Scholes) result due to a nonzero interest rate: if again X_t denotes the exponential martingale with parameter σ and S_0 the stock value at time 0, then

$$\mathbb{F}\left((S_T - K)^+\right) = \mathbb{E}\left((X_T - Ke^{-rT})^+\right).$$

This result will come out as a combined effect of an upward drift and a discount.

In the presence of bonds, a self-financing strategy is to be defined as a pair (a, b) of adapted processes with continuous paths such that the total portfolio value $C_t := a_t S_t + b_t \beta_t$ satisfies

$$dC_t = a_t dS_t + b_t d\beta_t.$$

Theorem 10.3 *On a stock market with stocks and bonds the fair price at time 0 of a claim to $g(S_T)$ at time $T > 0$ is*

$$\mathbb{F}(g(S_T)) = e^{-rT}\mathbb{E}(g(Y_T)),$$

where $(Y_t)_{t\in[0,T]}$ is the solution of the stochastic differential equation

$$dY_t = Y_t(rdt + \sigma dB_t) \qquad \text{with} \quad Y_0 = S_0.$$

Proof. Let $B := rx\frac{\partial}{\partial x} + \sigma^2 x^2 \frac{\partial^2}{\partial x^2}$ be the generator of the diffusion Y_t and define

$$f_t := e^{(T-t)(B-r)}g, \qquad (t \in [0,T]).$$

Let the strategy (a,b) be given by

$$a_t := f_t'(S_t) \quad \text{and} \quad b_t := \frac{1}{\beta_t}\Big(f_t(S_t) - S_t f_t'(S_t)\Big).$$

(Recall that $\beta_t = e^{rt}$ by (10.2).) Then $a_T S_T + b_T \beta_T = f_T(S_T) = g(S_T)$. Moreover, (a,b) is self-financing since

$$\begin{aligned} dC_t &:= df_t(S_t) \\ &= -(B-r)f_t(S_t)dt + f_t'(S_t)dS_t + \tfrac{1}{2}f_t''(S_t)(dS_t)^2 \\ &= \Big(-(B-r)f_t(S_t) + \tfrac{1}{2}\sigma^2 S_t^2 f_t''(S_t)\Big)dt + f_t'(S_t)dS_t \\ &= r\Big(f_t(S_t) - S_t f_t'(S_t)\Big)dt + f_t'(S_t)dS_t \\ &= b_t d\beta_t + a_t dS_t. \end{aligned}$$

It follows that

$$\mathbb{F}(g(S_T)) := f_0(S_0) = \left(e^{T(B-r)}g\right)(S_0) = \mathbb{E}\left(e^{-rT}g(Y_T)\right).$$

□

Corollary 10.4 *The fair price of a stock option is*

$$\begin{aligned} \mathbb{F}\left((S_T-K)^+\right) &= e^{-rT}\mathbb{E}\left((Y_T-K)^+\right) = \mathbb{E}\left((X_T - e^{-rT}K)^+\right) \\ &= S_0\Phi\left(\log\frac{S_0}{K} + (r+\tfrac{1}{2}\sigma^2)T\right) - Ke^{-rT}\Phi\left(\log\frac{S_0}{K} + (r-\tfrac{1}{2}\sigma^2)T\right). \end{aligned}$$

Proof. The first equality is Theorem 10.3, the second equality follows from the fact that $Y_t = e^{rt}X_t$, while the third equality is (10.4). □