Stochastic Differential Equations

Lecture notes for courses given
at Humboldt University Berlin
and University of Heidelberg

Markus Reiß
Institute of Applied Mathematics
University of Heidelberg
http://math.uni-heidelberg.de/studinfo/reiss

This version: February 12, 2007

Contents

Notation

The notation follows the usual conventions, nevertheless the general mathematical symbols that will be used are gathered in the first table. The notation of the different function spaces is presented in the second table. The last table shows some regularly used own notation.

General symbols

$A := B$	A is defined by B
$[a,b]$, (a,b)	closed, open interval from a to b
$\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$	$\{1,2,\dots\}$, $\{0,1,\dots\}$, $\{0,+1,-1,+2,-2,\dots\}$
$\mathbb{R}$, $\mathbb{R}^+$, $\mathbb{R}^-$, $\mathbb{C}$	$(-\infty,\infty)$, $[0,\infty)$, $(-\infty,0]$, complex numbers
$\mathrm{Re}(z), \mathrm{Im}(z), \bar{z}$	real part, imaginary part, complex conjugate of $z \in \mathbb{C}$
$\lfloor x \rfloor$	largest integer smaller or equal to $x \in \mathbb{R}$
$\lceil x \rceil$	smallest integer larger or equal to $x \in \mathbb{R}$
$a \vee b$, $a \wedge b$	maximum, minimum of a and b
$\|x\|$	modulus of $x \in \mathbb{R}$ or Euclidean norm of $x \in \mathbb{R}^d$
$A \subset B$	A is contained in B or $A = B$
$\mathrm{span}(v, w, \dots)$	the subspace spanned by $v, w, \dots$
$U + V$, $U \oplus V$	the sum, the direct sum ($U \cap V = \{0\}$) of U and V
$\dim V$, $\mathrm{codim}\, V$	linear dimension, codimension of V
$\mathrm{ran}\, T$, $\ker T$	range and kernel of the operator T
$\mathbb{E}_d$	identity matrix in $\mathbb{R}^{d \times d}$
$\det(M)$	determinant of M
$\\|T\\|$, $\\|T\\|_{X \to Y}$	operator norm of $T : X \to Y$
$f(\bullet)$, $g(\bullet_1, \bullet_2)$	the functions $x \mapsto f(x)$, $(x_1, x_2) \mapsto g(x_1, x_2)$
$\mathrm{supp}(f)$	support of the function f
$f\|_S$	function f restricted to the set S
$f', f'', f^{(m)}$	first, second, m-fold (weak) derivative of f
$f'(a+)$	derivative of f at a to the right
$\mathbf{1}_S$	indicator function of the set S
$\hat{f}$, $\mathcal{F}(f)$	$\hat{f}(\xi) = \mathcal{F}(f)(\xi) = \int_{\mathbb{R}} f(t) e^{-i\xi t} dt$ or estimator $\hat{f}$ of f
$\hat{a}$, $\mathcal{F}(a)$, $a \in M(I)$	$\hat{a}(\xi) = \mathcal{F}(a)(\xi) = \int_I e^{-i\xi t} da(t)$
log	natural logarithm
cos, sin, cosh, sinh	(hyperbolic) trigonometric functions

$\mathbb{P}$, $\mathbb{E}$, Var, Cov	probability, expected value, variance and covariance
$\mathcal{L}(X)$, $X \sim \mathbb{P}$	the law of X, $\mathcal{L}(X) = \mathbb{P}$
$X_n \stackrel{\mathbb{P}}{=} X$	X_n converges $\mathbb{P}$-stochastically to X
$X_n \stackrel{\mathcal{L}}{=} X$	X_n converges in law to X
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$\sigma(Z_i, i \in I)$	σ-algebra generated by $(Z_i)_{i \in I}$
δ_x	Dirac measure at x
$A \lesssim B$	$A = O(B)$, i.e. $\exists\, c > 0\, \forall p:\ A(p) \leq cB(p)$ (p parameter)
$A \gtrsim B$	$B \lesssim A$
$A \sim B$	$A \lesssim B$ and $B \lesssim A$

Function spaces and norms

$L^p(I, \mathbb{R}^d)$	p-integrable functions $f : I \to \mathbb{R}$ ($\int_I \lvert f\rvert^p < \infty$)
$C(I, \mathbb{R}^d)$	$\{f : I \to \mathbb{R}^d \mid f \text{ continuous}\}$
$C_K(I, \mathbb{R}^d)$	$\{f \in C(I, \mathbb{R}^d) \mid f \text{ has compact support}\}$
$C_0(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$	$\{f \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_2}) \mid \lim_{\lVert x\rVert \to \infty} f(x) = 0\}$
$\lVert f\rVert_\infty$	$\sup_x \lvert f(x)\rvert$
$\lVert \mu\rVert_{TV}$	total variation norm: $\lVert \mu\rVert_{TV} = \sup_{\lVert f\rVert_\infty = 1} \int f\, d\mu$

Specific definitions

$W(t)$	Brownian motion at time t
$X(t)$	solution process to SDE at time t
$(\mathcal{F}_t)$	filtration for $(W(t),\, t \geq 0)$

Chapter 1

Stochastic integration

1.1 White Noise

Many processes in nature involve random fluctuations which we have to account for in our models. In principle, everything can be random and the probabilistic structure of these random influences can be arbitrarily complicated. As it turns out, the so called "white noise" plays an outstanding role.

Engineers want the white noise process $(\dot{W}(t),\, t \in \mathbb{R})$ to have the following properties:

- The random variables $\{\dot{W}(t) \,|\, t \in \mathbb{R}\}$ are independent.
- $\dot{W}$ is stationary, that is the distribution of $(\dot{W}(t+t_1), \dot{W}(t+t_2), \ldots, \dot{W}(t+t_n))$ does not depend on t.
- The expectation $\mathbb{E}[\dot{W}(t)]$ is zero.

Hence, this process is supposed to model independent and identically distributed shocks with zero mean. Unfortunately, mathematicians can prove that such a real-valued stochastic process cannot have measurable trajectories $t \mapsto \dot{W}(t)$ except for the trivial process $\dot{W}(t) = 0$.

1.1.1 Problem. *If $(t, \omega) \mapsto \dot{W}(t, \omega)$ is jointly measurable with $\mathbb{E}[\dot{W}(t)^2] < \infty$ and $\dot{W}$ has the above stated properties, then for all $t \geq 0$*

$$\mathbb{E}\Big[\Big(\int_0^t \dot{W}(s)\, ds\Big)^2\Big] = 0$$

holds and $\dot{W}(t) = 0$ almost surely. Can we relax the hypothesis $\mathbb{E}[\dot{W}(t)^2] < \infty$?

Nevertheless, applications forced people to consider equations like

$$\dot{x}(t) = \alpha x(t) + \dot{W}(t), \quad t \geq 0.$$

The way out of this dilemma is found by looking at the corresponding integrated equation:

$$x(t) = x(0) + \int_0^t \alpha x(s)\, ds + \int_0^t \dot W(s)\, ds, \quad t \geq 0.$$

What properties should we thus require for the integral process $W(t) := \int_0^t \dot W(s) ds$, $t \geq 0$? A straight-forward deduction (from wrong premises...) yields

- $W(0) = 0$.
- The increments $(W(t_1) - W(t_2), W(t_3) - W(t_4), \ldots, W(t_{n-1}) - W(t_n))$ are independent for $t_1 \geq t_2 \geq \cdots \geq t_n$.
- The increments are stationary, that is $W(t_1 + t) - W(t_2 + t) \stackrel{\mathcal{L}}{=} W(t_1) - W(t_2)$ holds for all $t \geq 0$.
- The expectation $\mathbb{E}[W(t)]$ is zero.
- The trajectories $t \mapsto W(t)$ are continuous.

The last point is due to the fact that integrals over measurable (and integrable) functions are always continuous. It is highly nontrivial to show that – up to indistinguishability and up to the norming $\operatorname{Var}[W(1)] = 1$ – the only stochastic process fulfilling these properties is Brownian motion (also known as Wiener process) (Øksendal 1998). Recall that Brownian motion is almost surely nowhere differentiable!

Rephrasing the stochastic differential equation, we now look for a stochastic process $(X(t),\, t \geq 0)$ satisfying

$$X(t) = X(0) + \int_0^t \alpha X(s) ds + W(t), \quad t \geq 0, \tag{1.1.1}$$

where $(W(t),\, t \geq 0)$ is a standard Brownian motion. The precise formulation involving filtrations will be given later, here we shall focus on finding processes X solving (1.1.1).

The so-called variation of constants approach in ODEs would suggest the solution

$$X(t) = X(0)e^{\alpha t} + \int_0^t e^{\alpha(t-s)} \dot W(s)\, ds, \tag{1.1.2}$$

which we give a sense (in fact, that was Wiener's idea) by partial integration:

$$X(t) = X(0)e^{\alpha t} + W(t) + \int_0^t \alpha e^{\alpha(t-s)} W(s)\, ds. \tag{1.1.3}$$

This makes perfect sense now since Brownian motion is (almost surely) continuous and we could even take the Riemann integral. The verification that (1.1.3) defines a solution is straight forward:

$$\begin{aligned}\int_0^t \alpha X(s)\,ds &= X(0)\int_0^t \alpha e^{\alpha s}\,ds + \alpha\int_0^t W(s)\,ds + \alpha^2\int_0^t\int_0^s e^{\alpha(s-u)}W(u)\,du\,ds\\ &= X(0)(e^{\alpha t}-1) + \alpha\int_0^t W(s)\,ds + \alpha^2\int_0^t W(u)\int_u^t e^{\alpha(s-u)}\,ds\,du\\ &= X(0)(e^{\alpha t}-1) + \int_0^t \alpha W(u)e^{\alpha(t-u)}\,du\\ &= X(t) - X(0) - W(t).\end{aligned}$$

Note that the initial value $X(0)$ can be chosen arbitrarily. The expectation $\mu(t) := \mathbb{E}[X(t)] = \mathbb{E}[X(0)]e^{\alpha t}$ exists if $X(0)$ is integrable. Surprisingly this expectation function satisfies the deterministic linear equation, hence it converges to zero for $\alpha < 0$ and explodes for $\alpha > 0$. How about the variation around this mean value? Let us suppose that $X(0)$ is deterministic, $\alpha \neq 0$ and consider the variance function

$$\begin{aligned}v(t) := \operatorname{Var}[X(t)] &= \mathbb{E}\Big[\Big(W(t) + \int_0^t \alpha e^{\alpha(t-s)}W(s)\,ds\Big)^2\Big]\\ &= \mathbb{E}[W(t)^2] + 2\int_0^t \alpha e^{\alpha(t-s)}\,\mathbb{E}[W(t)W(s)]\,ds + \int_0^t\int_0^t \alpha^2 e^{\alpha(2t-u-s)}\,\mathbb{E}[W(s)W(u)]\,du\,ds\\ &= t + 2\int_0^t \alpha e^{\alpha(t-s)}s\,ds + 2\int_0^t\int_s^t \alpha^2 e^{\alpha(2t-u-s)}s\,du\,ds\\ &= t + \int_0^t \Big(2\alpha e^{\alpha(t-s)}s + 2\alpha(e^{2\alpha(t-s)} - e^{\alpha(t-s)})s\Big)\,ds\\ &= \tfrac{1}{2\alpha}\big(e^{2\alpha t}-1\big).\end{aligned}$$

This shows that for $\alpha < 0$ the variance converges to $\frac{1}{2|\alpha|}$ indicating a stationary behaviour, which will be made precise in the sequel. On the other hand, for $\alpha > 0$ we find that the standard deviation $\sqrt{v(t)}$ grows with the same order as $\mu(t)$ for $t \to \infty$ which lets us expect a very erratic behaviour.

In anticipation of the Itô calculus, the preceding calculation can be simplified by regarding (1.1.2) directly. The second moment of $\int_0^t e^{\alpha(t-s)}\,dW(s)$ is immediately seen to be $\int_0^t e^{2\alpha(t-s)}\,ds$, the above value.

1.1.2 Problem. *Justify the name "white noise" by calculating the expectation and the variance of the Fourier coefficients of $\dot{W}$ on $[0,1]$ by formal partial integration, i.e. using formally*

$$a_k = \int_0^1 \dot{W}(t)\sqrt{2}\sin(2\pi kt)\,dt = -\int_0^1 W(t)2\pi k\sqrt{2}\cos(2\pi kt)\,dt$$

and the analogon for the cosine coefficients. Conclude that the coefficients are i.i.d. standard normal, hence the intensity of each frequency component is equally strong ("white").

1.2 The Itô Integral

1.2.1 Construction in $\mathbf{L^2}$

We shall only need the Itô integral with respect to Brownian motion, so the general semimartingale theory will be left out. From now on we shall always be working on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where a filtration $(\mathcal{F}_t)_{t\geq 0}$, that is a nested family of σ-fields $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for $s \leq t$, is defined that satisfies the usual conditions:

- $\mathcal{F}_s = \bigcap_{t>s} \mathcal{F}_t$ for all $s \geq 0$ (right-continuity);
- all $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ are contained in $\mathcal{F}_0$.

A family $(X(t),\, t \geq 0)$ of $\mathbb{R}^d$-valued random variables on our probability space is called a stochastic process and this process is $(\mathcal{F}_t)$-adapted if all $X(t)$ are $\mathcal{F}_t$-measurable. Denoting the Borel σ-field on $[0, \infty)$ by $\mathcal{B}$, this process X is measurable if $(t, \omega) \mapsto X(t, \omega)$ is a $\mathcal{B} \otimes \mathcal{F}$-measurable mapping. We say that $(X(t),\, t \geq 0)$ is continuous if the trajectories $t \mapsto X(t, \omega)$ are continuous for all $\omega \in \Omega$. One can show that a process is measurable if it is (right-)continuous (Karatzas and Shreve 1991, Thm. 1.14).

1.2.1 Definition. *A (standard one-dimensional) Brownian motion with respect to the filtration $(\mathcal{F}_t)$ is a continuous $(\mathcal{F}_t)$-adapted real-valued process $(W(t),\, t \geq 0)$ such that*

- *W(0)=0;*
- *for all $0 \leq s \leq t$: $W(t) - W(s)$ is independent of $\mathcal{F}_s$;*
- *for all $0 \leq s \leq t$: $W(t) - W(s)$ is $\mathcal{N}(0, t-s)$-distributed.*

1.2.2 Remark. *Brownian motion can be constructed in different ways (Karatzas and Shreve 1991), but the proof of the existence of such a process is in any case non-trivial.*

We shall often consider a larger filtration $(\mathcal{F}_t)$ than the canonical filtration $(\mathcal{F}_t^W)$ of Brownian motion in order to include random initial conditions. Given a Brownian motion process W' on a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ with the canonical filtration $\mathcal{F}'_t = \sigma(W'(s),\, s \leq t)$ and the random variable X''_0 on a different space $(\Omega'', \mathcal{F}'', \mathbb{P}'')$, we can construct the product space with $\Omega = \Omega' \times \Omega''$, $\mathcal{F} = \mathcal{F}' \otimes \mathcal{F}''$, $\mathbb{P} = \mathbb{P}' \otimes \mathbb{P}''$ such that $W(t, \omega', \omega'') := W'(t, \omega')$ and $X_0(\omega', \omega'') := X''_0(\omega'')$ are independent and W is an $(\mathcal{F}_t)$-Brownian motion for $\mathcal{F}_t = \sigma(X_0; W(s), s \leq t)$. Note that X_0 is $\mathcal{F}_0$-measurable which always implies that X_0 and W are independent.

Our aim here is to construct the integral $\int_0^t Y(s)\, dW(s)$ with Brownian motion as integrator and a fairly general class of stochastic integrands Y.

1.2.3 Definition. *Let V be the class of real-valued stochastic processes $(Y(t),\, t \geq 0)$ that are adapted, measurable and that satisfy*

$$\|Y\|_V := \Big(\int_0^\infty \mathbb{E}\big[Y(t)^2\big]\, dt \Big)^{1/2} < \infty.$$

A process $Y \in V$ is called <u>simple</u> *if it is of the form*

$$Y(t,\omega) = \sum_{i=0}^{\infty} \eta_i(\omega) \mathbf{1}_{[t_i,t_{i+1})}(t),$$

with an increasing sequence $(t_i)_{i\geq 0}$ and $\mathcal{F}_{t_i}$-measurable random variables η_i.

For such simple processes $Y \in V$ we naturally define

$$\int_0^\infty Y(t)\,dW(t) := \sum_{i=0}^{\infty} \eta_i(W(t_{i+1}) - W(t_i)). \tag{1.2.1}$$

1.2.4 Proposition. *The right hand side in (1.2.1) converges in $L^2(\mathbb{P})$, hence the integral $\int_0^\infty Y(t)\,dW(t)$ is a $\mathbb{P}$-almost surely well defined random variable. Moreover the following isometry is valid for simple processes Y:*

$$\mathbb{E}\Big[\Big(\int_0^\infty Y(t)\,dW(t)\Big)^2\Big] = \|Y\|_V^2.$$

Proof. We show that the partial sums $S_k := \sum_{i=0}^k \eta_i(W(t_{i+1}) - W(t_i))$ form a Cauchy sequence in $L^2(\mathbb{P})$. Let $k \leq l$, then by the independence and zero mean property of Brownian increments we obtain

$$\begin{aligned}
\mathbb{E}\big[\big(S_l - S_k\big)^2\big] &= \sum_{i=k+1}^{l} \mathbb{E}\big[\big(\eta_i(W(t_{i+1}) - W(t_i))\big)^2\big] \\
&\quad + 2 \sum_{k+1\leq i<j\leq l} \mathbb{E}[\eta_i(W(t_{i+1}) - W(t_i))\eta_j]\,\mathbb{E}[W(t_{j+1}) - W(t_j)] \\
&= \sum_{i=k+1}^{l} \mathbb{E}\big[\big(\eta_i(W(t_{i+1}) - W(t_i))\big)^2\big] \\
&= \sum_{i=k+1}^{l} \mathbb{E}\big[\eta_i^2\big](t_{i+1} - t_i) \\
&= \int_{t_{k+1}}^{t_{l+1}} \mathbb{E}\big[Y(t)^2\big]\,dt.
\end{aligned}$$

Due to $\|Y\|_V < \infty$ the last line tends to zero for $k, l \to \infty$. By the completeness of $L^2(\mathbb{P})$ the Itô integral of Y is therefore well defined as the L^2-limit of (S_k). The same calculations as before also show

$$\mathbb{E}\big[S_k^2\big] = \int_0^{t_{k+1}} \mathbb{E}\big[Y(t)^2\big]\,dt.$$

By taking the limit $k \to \infty$ on both sides the asserted isometry property follows. □

The main idea for extending the Itô integral to general integrands in V is to show that the simple processes lie dense in V with respect to the $\|\bullet\|_V$-seminorm and to use the isometry to define the integral by approximation.

1.2.5 Proposition. *For any process $Y \in V$ there is a sequence of simple processes $(Y_n)_{n\geq 1}$ in V with $\lim_{n\to\infty}\|Y - Y_n\|_V = 0$.*

Proof. We proceed by relaxing the assumptions on Y step by step:

1. Y is continuous, $|Y(t)| \leq K$ for $t \leq T$ and $Y(t) = 0$ for $t \geq T$: Set $t_i^n := \frac{i}{n}$ and define
$$Y_n(t) := \sum_{i=0}^{Tn-1} Y(t_i)\mathbf{1}_{[t_i^n, t_{i+1}^n)}(t).$$
Then Y_n is clearly a simple process in V and by the continuity of $Y(t)$ the processes Y_n converge to Y pointwise for all (t, ω). Since $\|Y_n\|_V^2 \leq TK^2$ holds, the dominated convergence theorem implies $\lim_{n\to\infty}\|Y - Y_n\|_V = 0$.

2. $|Y(t)| \leq K$ for $t \leq T$ and $Y(t) = 0$ for $t \geq T$, $T \in \mathbb{N}$: Y can be approximated by continuous functions Y_n in V with these properties (only T replaced by $T + 1$). For this suppose that $h : [0, \infty) \to [0, \infty)$ is continuous, satisfies $h(t) = 0$ for $t \geq 1$ and $\int h = 1$. For $n \in \mathbb{N}$ define the convolution
$$Y_n(t) := \int_0^t Y(s)\tfrac{1}{n}h(n(t - s))\,ds.$$
Then Y_n is continuous, has support in $[0, T + \frac{1}{n}]$ and satisfies $|Y_n(t)| \leq K$ for all ω. Moreover, $Y_n(t)$ is $\mathcal{F}_t$-adapted so that $Y_n \in V$ holds. Real analysis shows that $\int(Y_n - Y)^2 \to 0$ holds for $n \to \infty$ and all ω, so the assertion follows again by dominated convergence.

3. $Y \in V$ arbitrary: The processes
$$Y_n(t) := \begin{cases} 0, & t \geq n \\ Y(t), & |Y(t)| \leq n, t < n \\ n, & Y(t) > n \\ -n, & Y(t) < -n \end{cases}$$
are as in the preceding step with $T = K = n$. Moreover, they converge to Y pointwise and satisfy $|Y_n(t, \omega)| \leq |Y(t, \omega)|$ for all (t, ω) so that dominated convergence gives $\lim_{n\to\infty}\|Y_n - Y\|_V = 0$.

Putting the different approximations together completes the proof. □

By the completeness of $L^2(\mathbb{P})$ and the isometry in Proposition 1.2.4 the following definition of the Itô integral makes sense, in particular it does not depend on the approximating sequence.

1.2.6 Definition. *For any $Y \in V$ choose a sequence (Y_n) of simple processes with $\lim_{n\to\infty} \|Y_n - Y\|_V = 0$ and define the Itô integral by*

$$\int_0^\infty Y(t)\,dW(t) := \lim_{n\to\infty} \int_0^\infty Y_n(t)\,dW(t),$$

where the limit is understood in an $L^2(\mathbb{P})$-sense.

For $0 \le A \le B$ and $Y \in V$ we set

$$\int_A^B Y(t)\,dW(t) = \int_0^\infty Y(t)\mathbf{1}_{[A,B]}(t)\,dW(t).$$

1.2.7 Problem.

1. *The quadratic covariation up to time t between two functions $f, g : \mathbb{R}^+ \to \mathbb{R}$ is given by*

$$\langle f, g\rangle_t = \lim_{|\Pi|\to 0} \sum_{t_i \in \Pi} (f(t_{i+1} \wedge t) - f(t_i \wedge t))(g(t_{i+1} \wedge t) - g(t_i \wedge t)) \quad \forall\, t \ge 0,$$

if the limit exists, where Π denotes a partition given by real numbers (t_i) with $t_0 = 0$, $t_i \uparrow \infty$ and width $|\Pi| = \max_i(t_{i+1} - t_i)$. We call $\langle f\rangle_t := \langle f, f\rangle_t$ the quadratic variation of f. Show that Brownian motion satisfies $\langle W\rangle_t = t$ for $t \ge 0$, when the involved limit is understood to hold in probability. Hint: consider convergence in $L^2(\mathbb{P})$.

2. *Show that the process X with $X(t) = W(t)\mathbf{1}_{[0,T]}(t)$ is in V for any $T \ge 0$. Prove the identity*

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

Hint: Consider $X^n = \sum_{k=0}^{n-1} W(\frac{kT}{n})\mathbf{1}_{[\frac{kT}{n}, \frac{(k+1)T}{n})}$ and use part 1.

1.2.2 Properties

In this subsection we gather the main properties of the Itô integral without giving proofs. Often the properties are trivial for simple integrands and follow by approximation for the general case, the continuity property will be shown in Corollary 1.2.11. Good references are Øksendal (1998) and Karatzas and Shreve (1991).

1.2.8 Theorem. *Let X and Y be processes in V then*

(a) $\mathbb{E}\Big[\Big(\int_0^\infty X(t)\,dW(t)\Big)^2\Big] = \|X\|_V^2$ *(Itô isometry)*

(b) $\mathbb{E}\Big[\int_0^\infty X(t)\,dW(t)\int_0^\infty Y(t)\,dW(t)\Big] = \int_0^\infty \mathbb{E}[X(t)Y(t)]\,dt$

(c) $\int_A^C X(t)\,dW(t) = \int_A^B X(t)\,dW(t) + \int_B^C X(t)\,dW(t)$ $\mathbb{P}$*-a.s. for all* $0 \le A \le B \le C$

(d) $\int_0^\infty (cX(t)+Y(t))\,dW(t) = c\int_0^\infty X(t)\,dW(t) + \int_0^\infty Y(t)\,dW(t)$ $\mathbb{P}$*-a.s. for all* $c \in \mathbb{R}$

(e) $\mathbb{E}\Big[\int_0^\infty X(t)\,dW(t)\Big] = 0$

(f) $\int_0^t X(s)\,dW(s)$ *is* $\mathcal{F}_t$*-measurable for* $t \ge 0$

(g) $\Big(\int_0^t X(s)\,dW(s),\, t \ge 0\Big)$ *is an* $\mathcal{F}_t$*-martingale*

(h) $\Big(\int_0^t X(s)\,dW(s),\, t \ge 0\Big)$ *has a continuous version*

(i) $\Big\langle \int_0^\bullet X(s)\,dW(s), \int_0^\bullet Y(s)\,dW(s)\Big\rangle_t = \int_0^t X(s)Y(s)\,ds$ *(quadratic covariation process)*

(j) $X(t)W(t) = \int_0^t X(s)\,dW(s) + \int_0^t W(s)\,dX(s)$ $\mathbb{P}$*-a.s. for X with bounded variation*

1.2.3 Doob's Martingale Inequality

1.2.9 Theorem. *Suppose $(X_n, \mathcal{F}_n)_{0\le n\le N}$ is a martingale. Then for every $p \ge 1$ and $\lambda > 0$*

$$\lambda^p\,\mathbb{P}\Big(\sup_{0\le n\le N}|X_n| \ge \lambda\Big) \le \mathbb{E}[|X_N|^p],$$

and for every $p > 1$

$$\mathbb{E}\Big[\sup_{0\le n\le N}|X_n|^p\Big] \le \Big(\frac{p}{p-1}\Big)^p\,\mathbb{E}[|X_N|^p].$$

Proof. Introduce the stopping time $\tau := \inf\{n \,|\, |X_n| \ge \lambda\} \wedge N$. Since $(|X_n|^p)$ is a submartingale the optional stopping theorem gives

$$\mathbb{E}[|X_N|^p] \ge \mathbb{E}[|X_\tau|^p] \ge \lambda^p\,\mathbb{P}\Big(\sup_n|X_n| \ge \lambda\Big) + \mathbb{E}[|X_N|^p\mathbf{1}_{\{\sup_n|X_n|<\lambda\}}],$$

which proves the first part. Moreover, we deduce from this inequality for any $K > 0$ and $p > 1$

$$\begin{aligned}\mathbb{E}\Big[\big(\sup_n |X_n| \wedge K\big)^p\Big] &= \mathbb{E}\Big[\int_0^K p\lambda^{p-1}\mathbf{1}_{\{\sup_n |X_n| \geq \lambda\}}\,d\lambda\Big] \\ &\leq \int_0^K p\lambda^{p-2}\,\mathbb{E}[|X_N|\mathbf{1}_{\{\sup_n |X_n| \geq \lambda\}}]\,d\lambda \\ &= p\,\mathbb{E}\Big[|X_N| \int_0^{\sup_n |X_n| \wedge K} \lambda^{p-2}\,d\lambda\Big] \\ &= \frac{p}{p-1}\,\mathbb{E}\big[|X_N|(\sup_n |X_n| \wedge K)^{p-1}\big].\end{aligned}$$

By Hölder's inequality,

$$\mathbb{E}\Big[\big(\sup_n |X_n| \wedge K\big)^p\Big] \leq \frac{p}{p-1}\,\mathbb{E}\Big[\big(\sup_n |X_n| \wedge K\big)^p\Big]^{(p-1)/p}\,\mathbb{E}[|X_N|^p]^{1/p},$$

which after cancellation and taking the limit $K \to \infty$ yields the asserted moment bound. □

1.2.10 Corollary. *(Doob's L^p-inequality) If $(X(t), \mathcal{F}_t)_{t \in I}$ is a right-continuous martingale indexed by a subinterval $I \subset \mathbb{R}$, then for any $p > 1$*

$$\mathbb{E}\Big[\sup_{t \in I} |X(t)|^p\Big]^{1/p} \leq \frac{p}{p-1}\,\sup_{t \in I}\,\mathbb{E}[|X(t)|^p]^{1/p}.$$

Proof. By the right-continuity of X we can restrict the supremum on the left to a countable subset $D \subset I$. This countable set D can be exhausted by an increasing sequence of finite sets $D_n \subset D$ with $\bigcup_n D_n = D$. Then the supremum over D_n increases monotonically to the supremum over D, the preceding theorem applies for each D_n and the monotone convergence theorem yields the asserted inequality. □

Be aware that Doob's L^p-inequality is different for $p = 1$ (Revuz and Yor 1999, p. 55).

1.2.11 Corollary. *For any $X \in V$ there exists a version of $\int_0^t X(s)\,dW(s)$ that is continuous in t, i.e. a continuous process $(J(t),\, t \geq 0)$ with*

$$\mathbb{P}\Big(J(t) = \int_0^t X(s)\,dW(s)\Big) = 1 \text{ for all } t \geq 0.$$

Proof. Let $(X_n)_{n \geq 1}$ be an approximating sequence for X of simple processes in V. Then by definition $I_n(t) := \int_0^t X_n(s)\,dW(s)$ is continuous in t for all ω. Moreover, $I_n(t)$ is an $\mathcal{F}_t$-martingale so that Doob's inequality and the Itô isometry yield the Cauchy property

$$\mathbb{E}\Big[\sup_{t \geq 0} |I_m(t) - I_n(t)|^2\Big] \leq 4\sup_{t \geq 0}\mathbb{E}[|I_m(t) - I_n(t)|^2] = 4\|X_m - X_n\|_V^2 \to 0$$

for $m, n \to \infty$. By the Chebyshev inequality and the Lemma of Borel-Cantelli there exist a subsequence $(I_{n_l})_{l\geq 1}$ and $L(\omega)$ such that $\mathbb{P}$-almost surely

$$\forall\, l \geq L(\omega) \ \sup_{t\geq 0} |I_{n_{l+1}}(t) - I_{n_l}(t)| \leq 2^{-l}.$$

Hence with probability one the sequence $(I_{n_l}(t))_{l\geq 1}$ converges uniformly and the limit function $J(t)$ is continuous. Since for all $t \geq 0$ the random variables $(I_{n_l}(t))_{l\geq 1}$ converge in probability to the integral $I(t) = \int_0^t X(s)\,dW(s)$, the random variables $I(t)$ and $J(t)$ must coincide for $\mathbb{P}$-almost all ω. □

In the sequel we shall consider only t-continuous versions of the stochastic integral.

1.2.4 Extension of the Itô integral

We extend the stochastic integral from processes in V to the more general class of processes V^*.

1.2.12 Definition. *Let V^* be the class of real-valued stochastic processes $(Y(t),\, t \geq 0)$ that are adapted, measurable and that satisfy*

$$\mathbb{P}\Big(\int_0^\infty Y(t)^2\,dt < \infty\Big) = 1.$$

1.2.13 Theorem. *For $Y \in V^*$ and $n \in \mathbb{N}$ consider the $\mathbb{R}^+ \cup \{+\infty\}$-valued stopping time (!)*

$$\tau_n(\omega) := \inf\Big\{T \geq 0 \,\Big|\, \int_0^T Y(t,\omega)^2\,dt \geq n\Big\}.$$

Then $\lim_{n\to\infty} \tau_n = \infty$ $\mathbb{P}$-a.s. and

$$\int_0^\infty Y(t)\,dW(t) := \lim_{n\to\infty} \int_0^{\tau_n} Y(t)\,dW(t)$$

exists as limit in probability. More precisely, we have $\mathbb{P}$-a.s.

$$\int_0^\infty Y(t)\,dW(t) = \int_0^{\tau_n} Y(t)\,dW(t) \ \text{ on } \Big\{\omega \,\Big|\, \int_0^\infty Y(t,\omega)^2\,dt < n\Big\}.$$

Proof. That $\tau_n = \infty$ holds for all $n \geq N$ on the event $\Omega_N := \{\omega \mid \int_0^\infty Y(t,\omega)^2\,dt < N\}$, is clear. By assumption the event $\bigcup_{n\geq 1} \Omega_n$ has probability one. Choosing $N \in \mathbb{N}$ so large that $\mathbb{P}(\bigcup_{n=1}^N \Omega_n) \geq 1 - \varepsilon$, the random variables $\int_0^{\tau_n} Y(t)\,dW(t)$ are constant for all $n \geq N$ with probability at least $1 - \varepsilon$. This implies that these random variables form a Cauchy sequence with respect to convergence in probability. By completeness the limit exists. The last assertion is obvious from the construction. □

1.2.14 Remark. *Observe that the first idea to set $\int Y(t)\,dW(t) = \int Y(t)\mathbf{1}_{\Omega_N}\,dW(t)$ for all $\omega \in \Omega_N$ is not feasible because $\mathbf{1}_{\Omega_N}$ is generally not adapted.*

By localisation via the stopping times (τ_n) one can infer the properties of the extended integral from Theorem 1.2.7. The last assertion of the following theorem is proved in (Revuz and Yor 1999, Prop. IV.2.13).

1.2.15 Theorem. *The stochastic integral over integrands in V^* has the same properties as that over integrands in V regarding linearity (Theorem 1.2.7(c,d)), measurability (1.2.7(f)) and existence of a continuous version (1.2.7(h)). However, it is only a local $(\mathcal{F}_t)$-martingale with quadratic covariation as in (1.2.7(i)).*

Moreover, if $Y \in V^*$ is left-continuous and Π is a partition of $[0,t]$, then the finite sum approximations converge in probability:

$$\int_0^t Y(s)\,dW(s) = \lim_{|\Pi|\to 0} \sum_{t_i\in\Pi} Y(t_i)(W(t_{i+1}) - W(t_i)).$$

1.2.5 The Fisk-Stratonovich integral

For integrands $Y \in V$ an alternative reasonable definition of the stochastic integral is by interpolation

$$\int_0^T Y(t)\circ dW(t) := \lim_{|\Pi|\to 0} \sum_{t_i\in\Pi} \tfrac{1}{2}(Y(t_{i+1})) + Y(t_i))(W(t_{i+1}) - W(t_i)),$$

where Π denotes a partition of $[0,T]$ with $0 = t_0 < t_1 < \dots < t_{n-1} = t_n = T$ and $|\Pi| = \max_i(t_{i+1} - t_i)$ and where the limit is understood in the $L^2(\Omega)$-sense. This is the Fisk-Stratonovich integral.

1.2.16 Theorem. *For an arbitrary integrand $Y \in V$ we have in probability*

$$\lim_{|\Pi|\to 0} \sum_{t_i\in\Pi} \tfrac{1}{2}(Y(t_{i+1}) + Y(t_i))(W(t_{i+1}) - W(t_i)) = \int_0^T Y(t)\,dW(t) + \tfrac{1}{2}\langle Y, W\rangle_T.$$

Proof. Since the process $Y^\Pi := \sum_{t_i\in\Pi} Y(t_i)\mathbf{1}_{[t_i,t_{i+1})}$ is a simple integrand in V and satisfies $\lim_{|\Pi|\to 0}\mathbb{E}[\|Y^\Pi - Y\|^2_{L^2(0,T)}] \to 0$, we have

$$\lim_{|\Pi|\to 0} \sum_{t_i\in\Pi} Y(t_i)(W(t_{i+1}) - W(t_i)) = \int_0^T Y(t)\,dW(t)$$

even in $L^2(\Omega)$ by Itô isometry. The assertion thus reduces to

$$\lim_{|\Pi|\to 0} \sum_{t_i\in\Pi} (Y(t_{i+1}) - Y(t_i))(W(t_{i+1}) - W(t_i)) = \langle Y, W\rangle_T,$$

which is just the definition of the quadratic covariation.between Y and W. □

1.2.17 Corollary. *The Fisk-Stratonovich integral is linear and has a continuous version, but it is usually not a martingale and not even centred.*

1.2.18 Example. *We have $\int_0^T W(t)\,dW(t) = \frac{1}{2}W(T)^2 - \frac{1}{2}T$, but $\int_0^T W(t)\circ dW(t) = \frac{1}{2}W(T)^2$.*

1.2.6 Multidimensional Case

1.2.19 Definition.

1. *An $\mathbb{R}^m$-valued $(\mathcal{F}_t)$-adapted stochastic process $W(t) = (W_1(t), \dots, W_m(t))^T$ is an m-dimensional Brownian motion if each component W_i, $i = 1, \dots, m$, is a one-dimensional $(\mathcal{F}_t)$-Brownian motion and all components are independent.*

2. *If Y is an $\mathbb{R}^{d\times m}$-valued stochastic process such that each component Y_{ij}, $1 \le i \le d$, $1 \le j \le m$, is an element of V^* then the multidimensional Itô integral $\int Y\,dW$ for m-dimensional Brownian motion W is an $\mathbb{R}^d$-valued random variable with components*

$$\Big(\int_0^\infty Y(t)\,dW(t)\Big)_i := \sum_{j=1}^m \int_0^\infty Y_{ij}(t)\,dW_j(t), \quad 1 \le i \le d.$$

1.2.20 Proposition. *The Itô isometry extends to the multidimensional case such that for $\mathbb{R}^{d\times m}$-valued processes X, Y with components in V and m-dimensional Brownian motion W*

$$\mathbb{E}\Big[\Big\langle \int_0^\infty X(t)\,dW(t), \int_0^\infty Y(t)\,dW(t)\Big\rangle\Big] = \int_0^\infty \sum_{i=1}^d \sum_{j=1}^m \mathbb{E}[X_{ij}(t)Y_{ij}(t)]\,dt.$$

Proof. The term in the brackets on the left hand side is equal to

$$\sum_{i=1}^d \sum_{j=1}^m \sum_{k=1}^m \int_0^\infty X_{ij}(t)\,dW_j(t) \int_0^\infty Y_{ik}(t)\,dW_k(t)$$

and the result follows from the one-dimensional Itô isometry once the following claim has been proved: stochastic integrals with respect to independent Brownian motions are uncorrelated (attention: they may well be dependent!).

For this let us consider two independent Brownian motions W_1 and W_2 and two simple processes Y_1, Y_2 in V on the same filtered probability space with

$$Y_k(t) = \sum_{i=0}^\infty \eta_{ik}(\omega)\mathbf{1}_{[t_i,t_{i+1})}(t), \quad k \in \{1,2\}.$$

The common partition of the time axis can always be achieved by taking a common refinement of the two partitions. Then by the $\mathcal{F}_{t_i}$-measurability of η_{ik} we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\int_0^\infty Y_1(t)\,dW_1(t)\int_0^\infty Y_2(t)\,dW_2(t)\Big]\\
&= \sum_{0\le i\le j<\infty} \mathbb{E}\big[\eta_{i1}\eta_{j2}(W_1(t_{i+1})-W_1(t_i))(W_2(t_{j+1})-W_2(t_j))\big]\\
&\quad + \sum_{0\le j<i<\infty} \mathbb{E}\big[\eta_{i1}\eta_{j2}(W_1(t_{i+1})-W_1(t_i))(W_2(t_{j+1})-W_2(t_j))\big]\\
&= \sum_{0\le i\le j<\infty} \mathbb{E}\big[\eta_{i1}\eta_{j2}(W_1(t_{i+1})-W_1(t_i))\big]\,\mathbb{E}\big[(W_2(t_{j+1})-W_2(t_j))\big]\\
&\quad + \sum_{0\le j<i<\infty} \mathbb{E}\big[\eta_{i1}\eta_{j2}(W_1(t_{j+1})-W_1(t_j))\big]\,\mathbb{E}\big[(W_2(t_{i+1})-W_2(t_i))\big]\\
&= 0
\end{aligned}$$

By Proposition 1.2.5 for each process in V there exists a sequence of simple processes such that the corresponding stochastic integrals converge in $L^2(\mathbb{P})$, which implies that the respective covariances converge, too. This density argument proves the general case. □

1.2.7 Itô's formula

For complete proofs see Karatzas and Shreve (1991) or any other textbook on stochastic integration. Note in particular that different proof strategies exist, e.g. Revuz and Yor (1999), and that many extensions exist.

1.2.21 Theorem. *For a process $h \in V^*$ and an adapted process $(g(t),\, t \ge 0)$ with $\int_0^T |g(t)|\,dt < \infty$ $\mathbb{P}$-almost surely for all $T > 0$ set*

$$X(t) := \int_0^t g(s)\,ds + \int_0^t h(s)\,dW(s), \quad t \ge 0.$$

Then $Y(t) = F(X(t), t)$, $t \ge 0$, with a function $F \in C^{2,1}(\mathbb{R}\times\mathbb{R}^+, \mathbb{R})$ satisfies

$$\begin{aligned}
Y(t) = Y(0) &+ \int_0^t \big(\tfrac{\partial F}{\partial t}(X(s),s) + \tfrac{\partial F}{\partial x}(X(s),s)g(s) + \tfrac{1}{2}\tfrac{\partial^2 F}{\partial x^2}(X(s),s)h^2(s)\big)\,ds\\
&+ \int_0^t \tfrac{\partial F}{\partial x}(X(s),s)h(s)\,dW(s), \quad t \ge 0.
\end{aligned}$$

Proof. We only sketch the proof and assume that F, $\frac{\partial F}{\partial x}$, $\frac{\partial^2 F}{\partial x^2}$, $\frac{\partial F}{\partial t}$ are even uniformly bounded. Then for a partition Π of $[0,t]$ with $0 = t_0 < t_1 < \dots < t_n = t$ we infer from Taylor's formula

$$\begin{aligned}F(t,X(t)) &= F(0,X(0)) + \sum_{k=1}^{n} F(t_k,X(t_k)) - F(t_{k-1},X(t_{k-1}))\\ &= F(0,X(0)) + \sum_{k=0}^{n-1}\Big(\tfrac{\partial F}{\partial t}\Delta t_k + \tfrac{\partial F}{\partial x}\Delta X(t_k)\\ &\quad + \tfrac{1}{2}\tfrac{\partial^2 F}{\partial x^2}(\Delta X(t_k))^2 + o(\Delta t_k) + O((\Delta t_k)(\Delta X(t_k))) + o((\Delta X(t_k))^2)\Big),\end{aligned}$$

where all derivatives are evaluated at $(X(t_k),t_k)$ and where we have set $\Delta t_k = t_{k+1}-t_k$, $\Delta X(t_k) = X(t_{k+1}) - X(t_k)$. If we now let the width of the partition $|\Pi|$ tend to zero, we obtain by the continuity of X and the construction of the Riemann integral

$$\sum_{k=0}^{n-1} \tfrac{\partial F}{\partial t}\Delta t_k \to \int_0^t \tfrac{\partial F}{\partial t}(X(s),s)\,ds$$

and by the identity $\Delta X(t_k) = \int_{t_k}^{t_{k+1}} g(s)\,ds + \int_{t_k}^{t_{k+1}} h(s)\,dW(s)$ and the construction of the Itô integral

$$\sum_{k=0}^{n-1} \tfrac{\partial F}{\partial x}\Delta X(t_k) \to \int_0^t \tfrac{\partial F}{\partial x}(X(s),s)g(s)\,ds + \int_0^t \tfrac{\partial F}{\partial x}(X(s),s)h(s)\,dW(s)$$

with convergence in $L^2(\mathbb{P})$. Note that for the precise derivation of the stochastic integral we have to consider as approximating integrands the processes

$$Y_\Pi(s) = h(s)\sum_{k=0}^{n-1} \tfrac{\partial F}{\partial x}(t_k,X(t_k))\mathbf{1}_{[t_k,t_{k+1})}(s), \quad s\in[0,t].$$

The third term converges to the quadratic variation process, using that an absolutely continuous function has zero quadratic variation:

$$\sum_{k=0}^{n-1} \tfrac{\partial^2 F}{\partial x^2}(\Delta X(t_k))^2 \to \int_0^t \tfrac{\partial^2 F}{\partial x^2}(X(s),s)h^2(s)\,ds.$$

The remainder terms converge to zero owing to the finite variation of $\int_0^\bullet g(s)\,ds$ and the finite quadratic variation of $\int_0^\bullet h(s)\,dW(s)$, which implies that the respective higher order variations vanish. □

1.2.22 Theorem. *For an $\mathbb{R}^{d\times m}$-valued process h with components in V^* and an adapted $\mathbb{R}^d$-valued process $(g(t),\, t\ge 0)$ with $\int_0^T \|g(t)\|\,dt < \infty$ $\mathbb{P}$-almost surely for all $T>0$ set*

$$X(t) := \int_0^t g(s)\,ds + \int_0^t h(s)\,dW(s), \quad t\ge 0,$$

where W is an m-dimensional Brownian motion. Then $Y(t) = F(X(t), t)$, $t \geq 0$, with a function $F \in C^{2,1}(\mathbb{R}^d \times \mathbb{R}^+, \mathbb{R}^p)$ satisfies

$$\begin{aligned} Y(t) = Y(0) + \int_0^t \Big(& \tfrac{\partial F}{\partial t}(s, X(s)) + DF(X(s), s) g(s) \\ & + \tfrac{1}{2} \sum_{i,j=1}^{d} \tfrac{\partial^2 F}{\partial x_i \partial x_j}(X(s), s) \Big(\sum_{l=1}^{m} h_{i,l}(s) h_{j,l}(s) \Big) \Big) \, ds \\ & + \int_0^t DF(X(s), s) h(s) \, dW(s), \quad t \geq 0. \end{aligned}$$

Here $DF = (\partial_{x_i} F_j)_{1 \leq i \leq d, 1 \leq j \leq p}$ denotes the Jacobian of F.

1.2.23 Remark. *Itô's formula is best remembered in differential form*

$$dF = F_t \, dt + F_x \, dX(t) + \tfrac{1}{2} F_{xx} \, d\langle X \rangle_t \ \text{(one-dimensional)}.$$

A rule of thumb for deriving also the multi-dimensional formula is to simplify the Taylor expansion by proceeding formally and then substituting $dt_i dt_j = dt_i dW_j(t) = 0$ and $dW_i(t) dW_j(t) = \delta_{i,j} dt$.

If the stochastic integrals on the right hand side in the two preceding theorems are interpreted in the Fisk-Stratonovich sense and if h is constant, then the terms involving second derivatives do not appear in the corresponding formulae:

$$dF = F_t \, dt + F_x g \, dt + h F_x \circ dW(t) \ \text{(one-dimensional)}.$$

Note, however, that for non-constant h we should write $dF = F_t \, dt + F_x g \, dt + F_x \circ h dW(t)$, where the last term is a Stratonovich integral with respect to the continuous martingale $\int_0^\bullet h(s) \, dW(s)$ instead of Brownian motion.

1.2.24 Problem.

1. *Consider again Problem 1.2.7 and evaluate $\int_0^t W(s) \, dW(s)$ by regarding $Y(t) = W(t)^2$.*

2. *Show that $X(t) = \exp(\sigma W(t) + (a - \frac{\sigma^2}{2})t)$, W a one-dimensional Brownian motion, satisfies the linear Itô stochastic differential equation*

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

   What would be the solution of the same equation in the Fisk-Stratonovitch interpretation?

3. *Suppose W is an m-dimensional Brownian motion, $m \geq 2$, started in $x \neq 0$. Consider the process $Y(t) = \|W(t)\|$ (Euclidean norm) and find an expression for the differential $dY(t)$, assuming that W does not hit zero.*

The Itô formula allows a rather simple proof of Lévy's martingale characterisation of Brownian motion.

1.2.25 Theorem. *Let* $(M(t), \mathcal{F}_t,\, t \geq 0)$ *be a continuous* $\mathbb{R}^m$*-valued local martingale with* $M(0) = 0$ *and cross-variations* $\langle M_k, M_l \rangle_t = \delta_{kl} t$ *for* $1 \leq k, l \leq d$ $\mathbb{P}$*-almost surely. Then* $(M(t),\, t \geq 0)$ *is an* m*-dimensional* $(\mathcal{F}_t)$*-Brownian motion.*

Proof. We only sketch the proof for $m = 1$ and proper martingales M, details can be found in (Karatzas and Shreve 1991, Thm. 3.16); in particular the integration theory for general semimartingales. In order to show that M has independent normally distributed increments, it suffices to show

$$\mathbb{E}[\exp(iu(M(t) - M(s)))\,|\,\mathcal{F}_s] = \exp(-u^2(t-s)/2), \quad u \in \mathbb{R},\, t \geq s \geq 0.$$

By Itô's formula for general continuous semimartingales applied to real and imaginary part separately we obtain

$$\exp(iuM(t)) = \exp(iuM(s)) + iu \int_s^t \exp(iuM(v))\,dM(v) - \tfrac{1}{2}u^2 \int_s^t \exp(iuM(v))\,dv.$$

Due to $|\exp(iuM(v))| = 1$ the stochastic integral is a martingale and the function $F(t) = \mathbb{E}[\exp(iu(M(t) - M(s))),|\,\mathcal{F}_s]$ satisfies

$$F(t) = 1 - \tfrac{1}{2}u^2 \int_s^t F(v)\,dv \quad \mathbb{P}\text{-a.s.}$$

This integral equation has the unique solution $F(t) = \exp(-u^2(t-s)/2)$. □

Chapter 2

Strong solutions of SDEs

2.1 The strong solution concept

The first definition of a solution of a stochastic differential equation reflects the interpretation that the solution process X at time t is determined by the equation and the exogenous input of the initial condition and the path of the Brownian motion up to time t. Mathematically, this is translated into a measurability condition on X_t or equivalently into the smallest reasonable choice of the filtration to which X should be adapted, see condition (a) below.

2.1.1 Definition. *A strong solution X of the stochastic differential equation*

$$dX(t) = b(X(t), t)\, dt + \sigma(X(t), t)\, dW(t), \quad t \geq 0, \tag{2.1.1}$$

with $b : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^{d\times m}$ measurable, on the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to the fixed m-dimensional Brownian motion W and the independent initial condition X_0 over this probability space is a stochastic process $(X(t),\, t \geq 0)$ satisfying:

(a) X is adapted to the filtration $(\mathcal{G}_t)$, where $\mathcal{G}_t^0 := \sigma(W(s),\, 0 \leq s \leq t) \vee \sigma(X_0)$ and $\mathcal{G}_t$ is the completion of $\bigcap_{s>t} \mathcal{G}_s^0$ with $\mathbb{P}$-null sets;

(b) X is a continuous process;

(c) $\mathbb{P}(X(0) = X_0) = 1$;

(d) $\mathbb{P}(\int_0^t \|b(X(s), s)\| + \|\sigma(X(s), s)\|^2\, ds < \infty) = 1$ holds for all $t > 0$;

(e) With probability one we have

$$X(t) = X(0) + \int_0^t b(X(s), s)\, ds + \int_0^t \sigma(X(s), s)\, dW(s), \quad \forall\, t \geq 0.$$

2.1.2 Remark. *It can be shown (Karatzas and Shreve 1991, Section 2.7) that the completion of the filtration of Brownian motion (or more generally of any strong Markov process) is right-continuous. This means that $\mathcal{G}_t$ equals already the completion of $\mathcal{G}_t^0$.*

With this definition at hand the notion of the existence of a strong solution is clear. We will say that strong uniqueness of a solution holds, only if the construction of a strong solution is unique on any probability space carrying the random elements W and X_0, where X_0 is an arbitrary initial condition.

2.1.3 Definition. *Suppose that, whenever $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space equipped with a Brownian motion W and an independent random variable X_0, any two strong solutions X and X' of (2.1.1) with initial condition X_0 satisfy $\mathbb{P}(\forall t \geq 0 : X_t = X'_t) = 1$. Then we say that strong uniqueness holds for equation (2.1.1) or more precisely for the pair (b, σ).*

2.1.4 Remark. *Since solution processes are by definition continuous and $\mathbb{R}^+$ is separable, it suffices to have the weaker condition $\mathbb{P}(X_t = X'_t) = 1$ for all $t \geq 0$ in the above definition.*

2.1.5 Problem. *Consider the stochastic differential equation*

$$dX(t) = -(1-t)^{-1} X(t)\,dt + dW(t), \quad 0 \leq t < 1.$$

Find a suitable modification of the notion of a strong solution for SDEs defined on bounded time intervals and check that the following process, called Brownian bridge, is a strong solution with initial condition $X(0) = 0$:

$$X(t) = (1-t) \int_0^t (1-s)^{-1}\,dW(s), \quad 0 \leq t < 1.$$

Does $\lim_{t \uparrow 1} X(t)$ exist in some sense of convergence?

2.2 Uniqueness

2.2.1 Example. *Consider the one-dimensional equation*

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

with a bounded, Borel-measurable function $b : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}$ that is nonincreasing in the first variable. Then strong uniqueness holds for this equation, that is for the pair $(b, 1)$. To prove this define for two strong solutions X and X' on the same filtered probability space the process $D(t) := X(t) - X'(t)$. This process is (weakly) differentiable with

$$\frac{d}{dt} D^2(t) = 2D(t)\dot{D}(t) = 2(X(t) - X'(t))(b(X(t), t) - b(X'(t), t)) \leq 0, \text{ a.e.}$$

From $X(0) = X'(0)$ we infer $D^2(t) = 0$ for all $t \geq 0$.

Already for deterministic differential equations examples of nonuniqueness are well known. For instance, the differential equation $\dot{x}(t) = |x(t)|^\alpha$ with $0 < \alpha < 1$ and $x(0) = 0$ has the family of solutions $x_\tau(t) = ((t-\tau)/\beta)^\beta$ for $t \geq \tau$, $x_\tau(t) = 0$ for $t \leq \tau$ with $\beta = 1/(1-\alpha)$ and $\tau \geq 0$. The usual sufficient condition for uniqueness in the deterministic theory is Lipschitz continuity of the analogue of the drift function in the space variable. Also for SDEs Lipschitz continuity, even in its local form, suffices. First, we recall the classical Gronwall Lemma.

2.2.2 Lemma. *Let $T > 0$, $c \geq 0$ and $u, v : [0,T] \to \mathbb{R}^+$ be measurable functions. If u is bounded and v is integrable, then*

$$u(t) \leq c + \int_0^t u(s)v(s)\,ds \quad \forall\, t \in [0,T]$$

implies

$$u(t) \leq c \exp\Big(\int_0^t v(s)\,ds\Big), \quad t \in [0,T].$$

Proof. Suppose $c > 0$ and set

$$z(t) := c + \int_0^t u(s)v(s)\,ds, \quad t \in [0,T].$$

Then $u(t) \leq z(t)$, $z(t)$ is weakly differentiable and for almost all t

$$\frac{\dot{z}(t)}{z(t)} = \frac{u(t)v(t)}{z(t)} \leq v(t)$$

holds so that $\log(z(t)) \leq \log(z(0)) + \int_0^t v(s)\,ds$ follows. This shows that

$$u(t) \leq z(t) \leq c \exp\Big(\int_0^t v(s)\,ds\Big), \quad t \in [0,T].$$

For $c = 0$ apply the inequality for $c_n > 0$ with $\lim_n c_n = 0$ and take the limit. □

2.2.3 Theorem. *Suppose that b and σ are locally Lipschitz continuous in the space variable, that is, for all $n \in \mathbb{N}$ there is a $K_n > 0$ such that for all $t \geq 0$ and all $x, y \in \mathbb{R}^d$ with $\|x\|, \|y\| \leq n$*

$$\|b(x,t) - b(y,t)\| + \|\sigma(x,t) - \sigma(y,t)\| \leq K_n \|x - y\|$$

holds. Then strong uniqueness holds for equation (2.1.1).

Proof. Let two solutions X and X' of (2.1.1) with the same initial condition X_0 be given on some common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define the stopping times $\tau_n := \inf\{t > 0 \,|\, \|X(t)\| \geq n\}$ and τ'_n in the same manner for X', $n \in \mathbb{N}$. Then

$\tau_n^* := \tau_n \wedge \tau_n'$ converges $\mathbb{P}$-almost surely to infinity. The difference $X(t\wedge\tau_n^*) - X'(t\wedge\tau_n^*)$ equals $\mathbb{P}$-almost surely

$$\int_0^{t\wedge\tau_n^*} (b(X(s),s) - b(X'(s),s))\,ds + \int_0^{t\wedge\tau_n^*} (\sigma(X(s),s) - \sigma(X'(s),s))\,dW(s).$$

We conclude by the Itô isometry and Cauchy-Schwarz inequality:

$$\begin{aligned}
&\mathbb{E}[\|X(t\wedge\tau_n^*) - X'(t\wedge\tau_n^*)\|^2]\\
&\le 2\,\mathbb{E}\Big[\Big(\int_0^{t\wedge\tau_n^*} \|b(X(s),s) - b(X'(s),s)\|\,ds\Big)^2\Big] + 2\,\mathbb{E}\Big[\int_0^{t\wedge\tau_n^*} \|\sigma(X(s),s) - \sigma(X'(s),s)\|^2\,ds\Big]\\
&\le 2TK_n^2 \int_0^t \mathbb{E}[\|X(s\wedge\tau_n^*) - X'(s\wedge\tau_n^*)\|^2]\,ds + 2K_n^2 \int_0^t \mathbb{E}[\|X(s\wedge\tau_n^*) - X'(s\wedge\tau_n^*)\|^2]\,ds.
\end{aligned}$$

By Gronwall's inequality we conclude $\mathbb{P}(X(t\wedge\tau_n^*) = X'(t\wedge\tau_n^*)) = 1$ for all $n\in\mathbb{N}$ and $t\in[0,T]$. Letting $n,T\to\infty$, we see that $X(t) = X'(t)$ holds $\mathbb{P}$-almost surely for all $t\ge 0$ and by Remark 2.1.4 strong uniqueness follows. □

2.2.4 Remark. *In the one-dimensional case strong uniqueness already holds for Hölder-continuous diffusion coefficient σ of order $1/2$, see (Karatzas and Shreve 1991, Proposition 5.2.13) for more details and refinements.*

2.3 Existence

In the deterministic theory differential equations are usually solved locally around the initial condition. In the stochastic framework one is rather interested in global solutions and then uses appropriate stopping in order to solve an equation up to some random explosion time. To exclude explosions in finite time, the linear growth of the coefficients suffices. The standard example for explosion is the ODE

$$\dot{x}(t) = x(t)^2, \quad t\ge 0, \qquad x(0)\neq 0.$$

Its solution is given by $x(t) = 1/(x_0^{-1} - t)$ which explodes for $x_0 > 0$ and $t\uparrow x_0^{-1}$. Note already here that with the opposite sign $\dot{x}(t) = -x(t)^2$ the solution $x(t) = x(0)/(1+t)$ exists globally. Intuitively, the different behaviour is clear because in the first case x grows the faster the further away from zero it is ("positive feedback"), while in the second case x monotonically converges to zero ("negative feedback").

We shall first establish an existence theorem under rather strong growth and Lipschitz conditions and then later improve on that.

2.3.1 Theorem. *Suppose that the coefficients satisfy the global Lipschitz and linear growth conditions*

$$\|b(x,t) - b(y,t)\| + \|\sigma(x,t) - \sigma(y,t)\| \le K\|x-y\| \quad \forall x,y\in\mathbb{R}^d,\ t\ge 0 \tag{2.3.1}$$

$$\|b(x,t)\| + \|\sigma(x,t)\| \le K(1+\|x\|) \quad \forall x\in\mathbb{R}^d,\ t\ge 0 \tag{2.3.2}$$

with some constant $K > 0$. Moreover, suppose that on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ there exists an m-dimensional Brownian motion W and an initial condition X_0 with $\mathbb{E}[\|X_0\|^2] < \infty$. Then there exists a strong solution of the SDE (2.1.1) with initial condition X_0 on this probability space, which in addition satisfies with some constant $C > 0$ the moment bound

$$\mathbb{E}[\|X(t)\|^2] \le C(1 + \mathbb{E}[\|X_0\|^2])e^{Ct^2}, \quad t \ge 0.$$

Proof. As in the deterministic case we perform successive approximations and apply a Banach fixed point argument ("Picard-Lindelöf iteration"). Define recursively

$$X^0(t) := X_0, \quad t \ge 0 \tag{2.3.3}$$

$$X^{n+1}(t) := X_0 + \int_0^t b(X^n(s), s)\, ds + \int_0^t \sigma(X^n(s), s)\, dW(s), \quad t \ge 0. \tag{2.3.4}$$

Obviously, the processes X^n are continuous and adapted to the filtration generated by X_0 and W. Let us fix some $T > 0$. We are going to show that for arbitrary $t \in [0, T]$

$$\mathbb{E}\Big[\sup_{0 \le s \le t} \|X^{n+1}(s) - X^n(s)\|^2\Big] \le C_1 \frac{(C_2 t)^n}{n!} \tag{2.3.5}$$

holds with suitable constants C_1, $C_2 > 0$ independent of t and n and $C_2 = O(T)$. Let us see how we can derive the theorem from this result. From Chebyshev's inequality we obtain

$$\mathbb{P}\Big(\sup_{0 \le s \le T} \|X^{n+1}(s) - X^n(s)\| > 2^{-n-1}\Big) \le 4C_1 \frac{(4C_2 T)^n}{n!}$$

The term on the right hand side is summable over n, whence by the Borel-Cantelli Lemma we conclude

$$\mathbb{P}\Big(\text{for infinitely many } n\text{: } \sup_{0 \le s \le T} \|X^{n+1}(s) - X^n(s)\| > 2^{-n-1}\Big) = 0.$$

Therefore, by summation $\sup_{m \ge 1} \sup_{0 \le s \le T} \|X^{n+m}(s) - X^n(s)\| \le 2^{-n}$ holds for all $n \ge N(\omega)$ with some $\mathbb{P}$-almost surely finite random index $N(\omega)$. In particular, the random variables $X^n(s)$ form a Cauchy sequence $\mathbb{P}$-almost surely and converge to some limit $X(s)$, $s \in [0, T]$. Obviously, this limiting process X does not depend on T and is thus defined on $\mathbb{R}^+$. Since the convergence is uniform over $s \in [0, T]$, the limiting process X is continuous. Of course, it is also adapted by the adaptedness of X^n. Taking the limit $n \to \infty$ in equation (2.3.4), we see that X solves the SDE (2.1.1) up to time T because of

$$\sup_{0 \le s \le T} \|b(X^n(s), s) - b(X_T(s), s)\| \le K \sup_{0 \le s \le T} \|X^n(s) - X_T(s)\| \to 0 \text{ (in } L^2(\mathbb{P}))$$

$$\mathbb{E}\Big[\|\sigma(X^n(\bullet), \bullet) - \sigma(X(\bullet), \bullet)\|^2_{V([0,T])}\Big] \le K^2 T \sup_{0 \le s \le T} \mathbb{E}[\|X_n(s) - X(s)\|^2] \to 0.$$

Since $T > 0$ was arbitrary, the equation (2.1.1) holds for all $t \geq 0$. From estimate (2.3.5) and the asymptotic bound $C_2 = O(T)$ we finally obtain by summation over n and putting $T = t$ the asserted estimate on $\mathbb{E}[\|X(t)\|^2]$.

It thus remains to establish the claimed estimate (2.3.5), which follows essentially from Doob's martingale inequality and the type of estimates used for proving Theorem 2.2.3. Proceeding inductively, we infer from the linear growth condition that (2.3.5) is true for $n = 0$ with some $C_1 > 0$. Assuming it to hold for $n - 1$, we obtain with a constant $D > 0$ from Doob's inequality:

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{0\le s\le t}\|X^{n+1}(s) - X^n(s)\|^2\Big] \\
&\le 2\,\mathbb{E}\Big[\sup_{0\le s\le t}\Big\|\int_0^s b(X^n(u),u) - b(X^{n-1}(u),u)\,du\Big\|^2\Big] \\
&\quad + 2\,\mathbb{E}\Big[\sup_{0\le s\le t}\Big\|\int_0^s \sigma(X^n(u),u) - \sigma(X^{n-1}(u),u)\,dW(u)\Big\|^2\Big] \\
&\le 2K^2 t\int_0^t \mathbb{E}[\|X^n(u) - X^{n-1}(u)\|^2]\,du + 2DK^2\int_0^t \mathbb{E}[\|X^n(u) - X^{n-1}(u)\|^2]\,du \\
&\le (2K^2TC_1 + 2DK^2)C_2^{n-1}\frac{t^n}{n!}.
\end{aligned}$$

The choice $C_2 = 2K^2(TC_1 + D)/C_1 = O(T)$ thus gives the result. □

The last theorem is the key existence theorem that allows generalisations into many directions. The most powerful one is essentially based on conditions such that a solution X exists locally and $\|X(t)\|^2$ remains bounded for all $t \geq 0$ ($L(x) = x^2$ is a Lyapunov function). Our presentation follows Durrett (1996).

2.3.2 Lemma. *Suppose X_1 and X_2 are adapted continuous processes with $X_1(0) = X_2(0)$ and $\mathbb{E}[\|X_1(0)\|^2] < \infty$. Let $\tau_R := \inf\{t \geq 0 \mid \|X_1(t)\| \geq R$ or $\|X_2(t)\| \geq R\}$. If both X_1 and X_2 satisfy the stochastic differential equation* (2.1.1) *on the random time interval $[0, \tau_R]$ with Lipschitz conditions on the coefficients b and σ, then $X_1(t\wedge\tau_R) = X_2(t \wedge \tau_R)$ holds $\mathbb{P}$-almost surely for all $t \geq 0$.*

Proof. We proceed as in the proof of inequality (2.3.5) and obtain for $0 \leq t \leq T$:

$$\begin{aligned}
\mathbb{E}\Big[\sup_{0\le s\le t\wedge\tau_R}\|X_1(s) - X_2(s)\|^2\Big] &\le 2K^2(t+D)\int_0^t \mathbb{E}[\|X_1(u\wedge\tau_R) - X_2(u\wedge\tau_R)\|^2]\,du \\
&\le 2K^2(T+D)\int_0^t \mathbb{E}\Big[\sup_{0\le s\le u\wedge\tau_R}\|X_1(s) - X_2(s)\|^2\Big]\,du.
\end{aligned}$$

Hence, Gronwall's Lemma implies that the expectation is zero and the result follows. □

2.3.3 Theorem. *Suppose the drift and diffusion coefficients b and σ are locally Lipschitz continuous in the space variable and satisfy for some $B \geq 0$*

$$2\langle x, b(x,t)\rangle + \operatorname{trace}(\sigma(x,t)\sigma(x,t)^T) \leq B(1+\|x\|^2), \quad \forall\, x \in \mathbb{R}^d,\, t \geq 0,$$

then the stochastic differential equation (2.1.1) *has a strong solution for any initial condition X_0 satisfying $\mathbb{E}[\|X_0\|^2] < \infty$.*

Proof. We extend the previous theorem by a suitable cut-off scheme. For any $R > 0$ define coefficient functions b_R, σ_R such that

$$b_R(x) = \begin{cases} b(x), & \|x\| \leq R, \\ 0, & \|x\| \geq 2R, \end{cases} \text{ and } \sigma_R(x) = \begin{cases} \sigma(x), & \|x\| \leq R, \\ 0, & \|x\| \geq 2R, \end{cases}$$

and b_R and σ_R are interpolated for $\|x\| \in (R, 2R)$ in such a way that they are Lipschitz continuous in the state variable. Then let X_R be the by Theorem 2.3.1 unique strong solution to the stochastic differential equation with coefficients b_R and σ_R. Introduce the stopping time $\tau_R := \inf\{t \geq 0 \,|\, \|X_R(t)\| \geq R\}$. Then by Lemma 2.3.2 $X_R(t)$ and $X_S(t)$ coincide for $t \leq \min(\tau_R, \tau_S)$ and we can define

$$X_\infty(t) := X_R(t) \text{ for } t \leq \tau_R.$$

The process X_∞ will be a strong solution of the stochastic differential equation (2.1.1) if we can show $\lim_{R\to\infty} \tau_R = \infty$ $\mathbb{P}$-almost surely.

Put $\varphi(x) = 1 + \|x\|^2$. Then Itô's formula yields for any $t, R > 0$

$$\begin{aligned}
&e^{-Bt}\varphi(X_R(t)) - \varphi(X_R(0)) \\
&= -B\int_0^t e^{-Bs}\varphi(X_R(s))\,ds + \sum_{i=1}^d \int_0^t e^{-Bs} 2X_{R,i}(s)\,dX_{R,i}(s) \\
&\quad + \frac{1}{2}\sum_{i=1}^d \int_0^t e^{-Bs} 2\sum_{j=1}^d \sigma_{ij}(X_R(s),s)^2\,ds \\
&= \text{local martingale} \\
&\quad + \int_0^t e^{-Bs}\Big(-B\varphi(X_R(s)) + 2\langle x, b_R(X_R(s),s)\rangle + \operatorname{trace}(\sigma_R(X_R(s),s)\sigma_R^T(X_R(s),s))\Big)\,ds.
\end{aligned}$$

Our assumption implies that $(e^{-B(t\wedge\tau_R)}\varphi(X_R(t\wedge\tau_R)))_{t\geq 0}$ is a supermartingale by the optional stopping theorem. We conclude

$$\begin{aligned}
\mathbb{E}[\varphi(X_0)] &\geq \mathbb{E}\big[e^{-B(t\wedge\tau_R)}\varphi(X_R(t\wedge\tau_R))\big] \\
&= \mathbb{E}\big[e^{-B(t\wedge\tau_R)}\varphi(X_\infty(t\wedge\tau_R))\big] \\
&\geq e^{-Bt}\,\mathbb{P}(\tau_R \leq t) \min_{\|x\|=R} \varphi(x).
\end{aligned}$$

Because of $\lim_{\|x\|\to\infty}\varphi(x) = \infty$ we have $\lim_{R\to\infty}\mathbb{P}(\tau_R \leq t) = 0$. Since the events $(\{\tau_R \leq t\})_{R>0}$ decrease, there exists for all $t > 0$ and $\mathbb{P}$-almost all ω an index R_0 such that $\tau_R(\omega) \geq t$ for all $R \geq R_0$, which is equivalent to $\tau_R \to \infty$ $\mathbb{P}$-almost surely. □

2.4 Explicit solutions

2.4.1 Linear Equations

In this paragraph we want to study the linear or affine equations

$$dX(t) = \big(A(t)X(t) + a(t)\big)\,dt + \sigma(t)\,dW(t), \quad t \geq 0. \tag{2.4.1}$$

Here, A is a $d \times d$-matrix, a is a d-dimensional vector and σ is a $d \times m$-dimensional matrix, where all objects are determinisic as well as measurable and locally bounded in the time variable. As usual, W is an m-dimensional Brownian motion and X a d-dimensional process.

The corresponding deterministic linear equation

$$\dot{x}(t) = A(t)x(t) + a(t), \quad t \geq 0, \tag{2.4.2}$$

has for every initial condition x_0 an absolutely continuous solution x, which is given by

$$x(t) = \Phi(t)\big(x_0 + \int_0^t \Phi^{-1}(s)a(s)\,ds\big), \quad t \geq 0,$$

where Φ is the so-called fundamental solution. This means that Φ solves the matrix equation

$$\dot{\Phi}(t) = A(t)\Phi(t), \quad t \geq 0, \qquad \text{with } \Phi(0) = \mathrm{Id}\,.$$

In the case of a matrix A that is constant in time, the fundamental solution is given by

$$\Phi(t) = e^{At} := \sum_{k=0}^{\infty} \frac{(tA)^k}{k!}.$$

2.4.1 Proposition. *The strong solution X of equation* (2.4.1) *with initial condition X_0 is given by*

$$X(t) = \Phi(t)\Big(X_0 + \int_0^t \Phi^{-1}(s)a(s)\,ds + \int_0^t \Phi^{-1}(s)\sigma(s)\,dW(s)\Big), \quad t \geq 0.$$

Proof. Apply Itô's formula. □

2.4.2 Problem.

1. *Show that the function $\mu(t) := \mathbb{E}[X(t)]$ under the hypothesis $\mathbb{E}[|X(0)|] < \infty$ satisfies the deterministic linear differential equation* (2.4.2).

2. *Assume that A, a and σ are constant. Calculate the covariance function* $\mathrm{Cov}(X(t), X(s))$ *and investigate under which conditions on A, a, σ and X_0 this function only depends on $|t-s|$ (weak stationarity). When do we have strong stationarity?*

2.4.2 Transformation methods

We follow the presentation by Kloeden and Platen (1992) and consider scalar equations that can be solved explicitly by suitable transformations.

Consider the scalar stochastic differential equation

$$dX(t) = \tfrac{1}{2}b(X(t))b'(X(t))\,dt + b(X(t))\,dW(t), \tag{2.4.3}$$

where $b : \mathbb{R} \to \mathbb{R}$ is continously differentiable and does not vanish and W is a one-dimensional Brownian motion. This equation is equivalent to the Fisk-Stratonovich equation

$$dX(t) = b(X(t)) \circ dW(t).$$

Define

$$h(x) := \int_c^x \frac{1}{b(y)}\,dy \text{ for some } c \in \mathbb{R}\,.$$

Then $X(t) := h^{-1}(W(t) + h(X_0))$, where h^{-1} denotes the inverse of h which exists by monotonicity, solves the equation (2.4.3). This follows easily from $(h^{-1})'(W(t) + h(X_0)) = b(X(t))$ and $(h^{-1})''(W(t) + h(X_0)) = b'(X(t))b(X(t))$.

2.4.3 Example.

1. *(geometric Brownian motion)* $dX(t) = \frac{\alpha^2}{2}X(t)\,dt + \alpha X(t)\,dW(t)$ *has the solution* $X(t) = X_0 \exp(\alpha W(t))$.

2. *The choice* $b(x) = \beta|x|^\alpha$ *for* $\alpha, \beta \in \mathbb{R}$ *corresponds formally to the equation*

$$dX(t) = \tfrac{1}{2}\alpha\beta^2|X(t)|^{2\alpha-1}\operatorname{sgn}(X(t))\,dt + \beta|X(t)|^\alpha\,dW(t).$$

For $\alpha < 1$ *we obtain formally the solution*

$$X(t) = \Big|\beta(1-\alpha)W(t) + |X_0|^{1-\alpha}\operatorname{sgn}(X_0)\Big|^{1/(1-\alpha)} \operatorname{sgn}(\beta(1-\alpha)W(t) + |X_0|^{1-\alpha}\operatorname{sgn}(X_0)).$$

This is well defined and indeed a strong solution if $\frac{1}{1-\alpha}$ *is nonnegative. The specific choice* $\alpha = \frac{n-1}{n}$ *with* $n \in \mathbb{N}$ *odd gives*

$$X(t) = (\beta n^{-1}W(t) + \sqrt[n]{X_0})^n.$$

For even n *and* $X_0 \geq 0$ *this formula defines a solution of*

$$dX(t) = \frac{(n-1)\beta^2}{n}X(t)^{(n-2)/n}\,dt + \beta X(t)^{(n-1)/n}\,dW(t),$$

and X *remains nonnegative for all times* $t \geq 0$. *Observe that a solution exists, although the coefficients are not locally Lipschitz. One can show that for* $n = 2$ *strong uniqueness holds, whereas for* $n > 2$ *also the trivial process* $X(t) = 0$ *is a solution.*

3. *The equation*

$$dX(t) = -a^2 \sin(X(t)) \cos^3(X(t))\, dt + a \cos^2(X(t))\, dW(t)$$

has for $X_0 \in (-\frac{\pi}{2}, \frac{\pi}{2})$ *the solution* $X(t) = \arctan(aW(t) + \tan(X_0))$, *which remains contained in the interval* $(-\frac{\pi}{2}, \frac{\pi}{2})$. *This can be explained by the fact that for* $x = \pm\frac{\pi}{2}$ *the coefficients vanish and for values* x *close to this boundary the drift pushes the process towards zero more strongly than the diffusion part can possibly disturb.*

4. *The equation*

$$dX(t) = a^2 X(t)(1 + X(t)^2)\, dt + a(1 + X(t)^2)\, dW(t)$$

is solved by $X(t) = \tan(aW(t) + \arctan X_0)$ *and thus explodes* $\mathbb{P}$*-almost surely in finite time.*

The transformation idea allows certain generalisations. With the same assumptions on b and the same definition of h we can solve the equation

$$dX(t) = \Big(\alpha b(X(t)) + \frac{1}{2} b(X(t)) b'(X(t))\Big)\, dt + b(X(t))\, dW(t)$$

by $X(t) = h^{-1}(\alpha t + W(t) + h(X_0))$. Equations of the type

$$dX(t) = \Big(\alpha h(X(t)) b(X(t)) + \frac{1}{2} b(X(t)) b'(X(t))\Big)\, dt + b(X(t))\, dW(t)$$

are solved by $X(t) = h^{-1}(e^{\alpha t} h(X_0) + e^{\alpha t} \int_0^t e^{-\alpha s}\, dW(s))$.

Finally, we consider for $n \in \mathbb{N}$, $n \geq 2$, the equation

$$dX(t) = (aX(t)^n + bX(t))\, dt + cX(t)\, dW(t).$$

Writing $Y(t) = X(t)^{1-n}$ we obtain

$$\begin{aligned} dY(t) &= (1-n)X(t)^{-n}\, dX(t) + \tfrac{1}{2}(1-n)(-n)X(t)^{-n-1} c^2 X^2(t)\, dt \\ &= (1-n)(a + (b - \tfrac{c^2}{2} n)Y(t))dt + (1-n)cY(t)\, dW(t). \end{aligned}$$

Hence, Y is a geometric Brownian motion and we obtain after transformation for all $X_0 \neq 0$

$$X(t) = e^{(b - \frac{c^2}{2})t + cW(t)} \Big(X_0^{1-n} + a(1-n) \int_0^t e^{(n-1)(b - \frac{c^2}{2})s + c(n-1)W(s)}\, ds \Big)^{1/(1-n)}.$$

In addition to the trivial solution $X(t) = 0$ we therefore always have a nonnegative global solution in the case $X_0 \geq 0$ and $a \leq 0$. For odd integers n and $a \leq 0$ a global solution exists for any initial condition, cf. Theorem 2.3.3. In the other cases it is easily seen that the solution explodes in finite time.

Chapter 3

Weak solutions of SDEs

3.1 The weak solution concept

We start with the famous example of H. Tanaka. Consider the scalar SDE

$$dX(t) = \operatorname{sgn}(X(t))\, dW(t), \quad t \geq 0, \qquad X(0) = 0, \tag{3.1.1}$$

where $\operatorname{sgn}(x) = \mathbf{1}_{(0,\infty)}(x) - \mathbf{1}_{(-\infty,0]}(x)$. Any adapted process X satisfying (3.1.1) is a continuous martingale with quadratic variation $\langle X \rangle_t = t$. Lévy's Theorem 1.2.25 implies that X has the law of Brownian motion. If X satisfies this equation, then so does $-X$, since the Lebesgue measure of $\{t \in [0,T] \,|\, X(t) = 0\}$ vanishes almost surely for any Brownian motion. Hence strong uniqueness cannot hold.

We now invert the roles of X and W, for equation (3.1.1) obviously implies $dW(t) = \operatorname{sgn}(X(t))\, dX(t)$. Hence, we take a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a Brownian motion X and consider the filtration $(\mathcal{F}_t^X)_{t\geq 0}$ generated by X and completed under $\mathbb{P}$. Then we define the process

$$W(t) := \int_0^t \operatorname{sgn}(X(s))\, dX(s), \quad t \geq 0.$$

W is a continuous $(\mathcal{F}_t^X)$-adapted martingale with quadratic variation $\langle W \rangle_t = t$, hence also an $(\mathcal{F}_t^X)$-Brownian motion. The couple (X, W) then solves the Tanaka equation. However, X is not a strong solution because the filtration $(\mathcal{F}_t^W)_{t\geq 0}$ generated by W and completed under $\mathbb{P}$ satisfies $\mathcal{F}_t^W \subsetneqq \mathcal{F}_t^X$ as we shall see.

For the proof let us take a sequence (f_n) of continuously differentiable functions on the real line that satisfy $f_n(x) = \operatorname{sgn}(x)$ for $|x| \geq \frac{1}{n}$ and $|f_n(x)| \leq 1$, $f_n(-x) = -f_n(x)$ for all $x \in \mathbb{R}$. If we set $F_n(x) = \int_0^x f_n(y)\, dy$, then $F_n \in C^2(\mathbb{R})$ and $\lim_{n\to\infty} F_n(x) = |x|$ holds uniformly on compact intervals. By Itô's formula for any solution X of (3.1.1)

$$F_n(X(t)) - \int_0^t f_n(X(s))dX(s) = \frac{1}{2}\int_0^t f_n'(X(s))ds, \quad t \geq 0,$$

follows and by Lebesgue's Theorem the left hand side converges in probability for $n \to \infty$ to $|X(t)| - \int_0^t \operatorname{sgn}(X(s))dX(s) = |X(t)| - W(t)$. By symmetry, $f_n'(x) = f_n'(|x|)$ and we have for $t \geq 0$ $\mathbb{P}$-almost surely

$$W(t) = |X(t)| - \lim_{n\to\infty} \frac{1}{2} \int_0^t f_n'(|X(s)|)ds.$$

Hence, $\mathcal{F}_t^W \subseteqq \mathcal{F}_t^{|X|}$ holds with obvious notation. The event $\{X(t) > 0\}$ has probability $\frac{1}{2} > 0$ and is not $\mathcal{F}_t^{|X|}$-measurable. Therefore $\mathcal{F}_t^X \setminus \mathcal{F}_t^{|X|}$ is non-void and $\mathcal{F}_t^W \subsetneqq \mathcal{F}_t^X$ holds for any solution X, which is thus not a strong solution in our definition. Note that the above derivation would be clearer with the aid of Tanaka's formula and the concept of local time.

3.1.1 Definition. *A weak solution of the stochastic differential equation* (2.1.1) *is a triple* (X, W), $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t\geq 0}$ *where*

(a) $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space equipped with the filtration* $(\mathcal{F}_t)_{t\geq 0}$ *that satisfies the usual conditions;*

(b) X *is a continuous,* $(\mathcal{F}_t)$*-adapted* $\mathbb{R}^d$*-valued process and* W *is an* m*-dimensional* $(\mathcal{F}_t)$*-Brownian motion on the probability space;*

(c) conditions (d) and (e) of Definition 2.1.1 are fulfilled.

The distribution $\mathbb{P}^{X(0)}$ *of* $X(0)$ *is called initial distribution of the solution* X.

3.1.2 Remark. *Any strong solution is also a weak solution with the additional filtration property* $\mathcal{F}_t^X \subseteqq \mathcal{F}_t^W \vee \sigma(X(0))$. *The Tanaka equation provides a typical example of a weakly solvable SDE that has no strong solution.*

3.1.3 Definition. *We say that pathwise uniqueness for equation* (2.1.1) *holds whenever two weak solutions* (X, W), $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t\geq 0}$ *and* (X', W), $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t')_{t\geq 0}$ *on a common probability space with a common Brownian motion with respect to both filtrations* $(\mathcal{F}_t)$ *and* $(\mathcal{F}_t')$, *and with* $\mathbb{P}(X(0) = X'(0)) = 1$ *satisfy* $\mathbb{P}(\forall\, t \geq 0 : \; X(t) = X'(t)) = 1$.

3.1.4 Definition. *We say that uniqueness in law holds for equation* (2.1.1) *whenever two weak solutions* (X, W), $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t\geq 0}$ *and* (X', W'), $(\Omega', \mathcal{F}', \mathbb{P}')$, $(\mathcal{F}_t')_{t\geq 0}$ *with the same initial distribution have the same law, that is* $\mathbb{P}(X(t_1) \in B_1, \ldots, X(t_n) \in B_n) = \mathbb{P}'(X'(t_1) \in B_1, \ldots, X'(t_n) \in B_n)$ *holds for all* $n \in \mathbb{N}$, $t_1, \ldots, t_n > 0$ *and Borel sets* $B_1, \ldots, B_n$.

3.1.5 Example. *For the Tanaka equation pathwise uniqueness fails because* X *and* $-X$ *are at the same time solutions. We have, however, seen that* X *must have the law of a Brownian motion and thus uniqueness in law holds.*

3.2 The two concepts of uniqueness

Let us discuss the notion of pathwise uniqueness and of uniqueness in law in some detail. When we consider weak solutions we are mostly interested in the law of the solution process so that uniqueness in law is usually all we require. However, as we shall see, the concept of pathwise uniqueness is stronger than that of uniqueness in law and if we reconsider the proof of Theorem 2.2.3 we immediately see that we have not used the special filtration properties of strong uniqueness and we obtain:

3.2.1 Theorem. *Suppose that b and σ are locally Lipschitz continuous in the space variable, that is, for all $n \in \mathbb{N}$ there is a $K_n > 0$ such that for all $t \geq 0$ and all $x, y \in \mathbb{R}^d$ with $\|x\|, \|y\| \leq n$*

$$\|b(x,t) - b(y,t)\| + \|\sigma(x,t) - \sigma(y,t)\| \leq K_n \|x - y\|$$

holds. Then pathwise uniqueness holds for equation (2.1.1).

The same remark applies to Example 2.2.1. As Tanaka's example has shown, pathwise uniqueness can fail when uniqueness in law holds. It is not clear, though, that the converse implication is true.

3.2.2 Theorem. *Pathwise uniqueness implies uniqueness in law.*

Proof. We have to show that two weak solutions (X_i, W_i), $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$, $(\mathcal{F}_t^i)$, $i = 1, 2$ on possibly different filtered probability spaces agree in distribution. The main idea is to define two weak solutions with the same law on a common space with the same Brownian motion and to apply the pathwise uniqueness assumption. To this end we set

$$S := \mathbb{R}^d \times C(\mathbb{R}^+, \mathbb{R}^m) \times C(\mathbb{R}^+, \mathbb{R}^d), \qquad \mathcal{S} = \text{Borel } \sigma\text{-field of } S$$

and consider the image measures

$$\mathbb{Q}_i(A) := \mathbb{P}_i((X_i(0), W_i, X_i) \in A), \quad A \in \mathcal{S}, \quad i = 1, 2.$$

Since $X_i(t)$ is by definition $\mathcal{F}_t^i$-measurable, $X_i(0)$ is independent of W_i under $\mathbb{P}_i$. If we call μ the law of $X_i(0)$ under $\mathbb{P}_i$ (which by assumption does not depend on i), we thus have that the product measure $\mu \otimes \mathbb{W}$ is the law of the first two coordinates $(X_i(0), W_i)$ under $\mathbb{P}_i$, where $\mathbb{W}$ denotes the Wiener measure. Since $C(\mathbb{R}^+, \mathbb{R}^k)$ is a Polish space, a regular conditional distribution (Markov kernel) K_i of X_i under $\mathbb{P}_i$ given $(X_i(0), W_i)$ exists (Karatzas and Shreve 1991, Section 5.3D) and we may write for Borel sets $F \subset \mathbb{R}^d \times C(\mathbb{R}^+, \mathbb{R}^m)$, $G \subset C(\mathbb{R}^+, \mathbb{R}^d)$

$$\mathbb{Q}_i(F \times G) = \int_F K_i(x_0, w; G)\, \mu(dx_0)\; \mathbb{W}(dw).$$

Let us now define

$$T = S \times C(\mathbb{R}^+, \mathbb{R}^d), \qquad \mathcal{T} = \text{Borel } \sigma\text{-field of } T$$

and equip this space with the probability measure

$$\mathbb{Q}(d(x_0, w, y_1, y_2)) = K_1(x_0, w; dy_1) K_2(x_0, w; dy_2) \mu(dx_0)\, \mathbb{W}(dw).$$

Finally, denote by $\mathcal{T}^*$ the completion of $\mathcal{T}$ under $\mathbb{Q}$ and consider the filtrations

$$\mathcal{T}_t = \sigma((x_0, w(s), y_1(s), y_2(s)),\ s \le t)$$

and its $\mathbb{Q}$-completion $\mathcal{T}_t^*$ and its right-continuous version $\mathcal{T}^{**} = \bigcap_{s>t} \mathcal{T}_s^*$. Then the projection on the first coordinate has under $\mathbb{Q}$ the law of the initial distribution of X_i and the projection on the second coordinate is under $\mathbb{Q}$ an $\mathcal{T}_t^{**}$-Brownian motion (recall Remark 2.1.2). Moreover, the distribution of the projection (w, y_i) under $\mathbb{Q}$ is the same as that of (W_i, X_i) under $\mathbb{P}_i$ such that we have constructed two weak solutions on the same probability space with the same initial condition and the same Brownian motion.

Pathwise uniqueness now implies $\mathbb{Q}(\{(x_0, w, y_1, y_2) \in T \,|\, y_1 = y_2\}) = 1$. This entails

$$\mathbb{P}_1((W_1, X_1) \in A) = \mathbb{Q}((w, y_1) \in A) = \mathbb{Q}((w, y_2) \in A) = \mathbb{P}_2((W_2, X_2) \in A).$$

□

The same methodology allows to prove the following, at a first glance rather striking result.

3.2.3 Theorem. *The existence of a weak solution and pathwise uniqueness imply the existence of a strong solution on any sufficiently rich probability space.*

Proof. See (Karatzas and Shreve 1991, Cor. 5.3.23). □

3.3 Existence via Girsanov's theorem

The Girsanov theorem is one of the main tools of stochastic analysis. In the theory of stochastic differential equations it often allows to extend results for a particular equation to those with more general drift coefficients. Abstractly seen, a Radon-Nikodym density for a new measure is obtained, under which the original process behaves differently. We only work in dimension one and start with a lemma on conditional Radon-Nikodym densities.

3.3.1 Lemma. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{H} \subset \mathcal{F}$ be a sub-σ-algebra and $f \in L^1(\mathbb{P})$ be a density, that is nonnegative and integrating to one. Then a new probability measure $\mathbb{Q}$ on $\mathcal{F}$ is defined by $\mathbb{Q}(d\omega) = f(\omega)\, \mathbb{P}(d\omega)$ and for any $\mathcal{F}$-measurable random variable X with $\mathbb{E}_{\mathbb{Q}}[|X|] < \infty$ we obtain*

$$\mathbb{E}_{\mathbb{Q}}[X \,|\, \mathcal{H}]\, \mathbb{E}_{\mathbb{P}}[f \,|\, \mathcal{H}] = \mathbb{E}_{\mathbb{P}}[Xf \,|\, \mathcal{H}] \quad \mathbb{P}\text{-}a.s.$$

3.3.2 Remark. *In the unconditional case we obviously have*

$$\mathbb{E}_{\mathbb{Q}}[X] = \int X \, d\,\mathbb{Q} = \int Xf \, d\,\mathbb{P} = \mathbb{E}_{\mathbb{P}}[Xf].$$

Proof. We show that the left-hand side is a version of the conditional expectation on the right. Since it is obviously $\mathcal{H}$-measurable, it suffices to verify

$$\int_H \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{H}] \, \mathbb{E}_{\mathbb{P}}[f \mid \mathcal{H}] \, d\,\mathbb{P} = \int_H Xf \, d\,\mathbb{P} = \int_H X \, d\,\mathbb{Q} \quad \forall\, H \in \mathcal{H}.$$

By the projection property of conditional expectations we obtain

$$\mathbb{E}_{\mathbb{P}}[\mathbf{1}_H \, \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{H}] \, \mathbb{E}_{\mathbb{P}}[f \mid \mathcal{H}]] = \mathbb{E}_{\mathbb{P}}[\mathbf{1}_H \, \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{H}] f] = \mathbb{E}_{\mathbb{Q}}[\mathbf{1}_H \, \mathbb{E}_{\mathbb{Q}}[X \mid \mathcal{H}]] = \mathbb{E}_{\mathbb{Q}}[\mathbf{1}_H X],$$

which is the above identity. □

3.3.3 Lemma. *Let $(\beta(t),\, 0 \le t \le T)$ be an $(\mathcal{F}_t)$-adapted process with $\beta \mathbf{1}_{t \le T} \in V^*$. Then*

$$M(t) := \exp\Big(-\int_0^t \beta(s) \, dW(s) - \frac{1}{2} \int_0^t \beta^2(s) \, ds\Big), \quad 0 \le t \le T,$$

is an $(\mathcal{F}_t)$-supermartingale. It is a martingale if and only if $\mathbb{E}[M(T)] = 1$ holds.

Proof. If we apply Itô's formula to M, we obtain

$$dM(t) = -\beta(t) M(t) \, dW(t), \quad 0 \le t \le T.$$

Hence, M is always a nonnegative local $\mathbb{P}$-martingale. By Fatou's lemma for conditional expectations we infer that M is a supermartingale and a proper martingale if and only if $\mathbb{E}_{\mathbb{P}}[M(T)] = \mathbb{E}_{\mathbb{P}}[M(0)] = 1$. □

3.3.4 Lemma. *M is a martingale if β satisfies one of the following conditions:*

1. *β is uniformly bounded;*

2. *Novikov's condition:*

$$\mathbb{E}\Big[\exp\Big(\frac{1}{2} \int_0^T \beta^2(t) \, dt\Big)\Big] < \infty;$$

3. *Kazamaki's condition:*

$$\mathbb{E}\Big[\exp\Big(\frac{1}{2} \int_0^T \beta(t) \, dW(t)\Big)\Big] < \infty.$$

Proof. By the previous proof we know that M solves the linear SDE $dM(t) = -\beta(t)M(t)\,dW(t)$ with $M(0) = 1$. Since $\beta(t)$ is uniformly bounded, the diffusion coefficient satisfies the linear growth and Lipschitz conditions and we could modify Theorem 2.3.1 to cover also stochastic coefficients and obtain equally that $\sup_{0\le t\le T}\mathbb{E}[M(t)^2]$ is finite. This implies $\beta M\mathbf{1}_{[0,T]} \in V$ and M is a martingale.

Alternatively, we prove $\beta M\mathbf{1}_{[0,T]} \in V$ by hand: If β is uniformly bounded by some $K > 0$, then we have for any $p > 0$ and any partition $0 = t_0 \le t_1 \le \cdots \le t_n = t$

$$
\begin{aligned}
&\mathbb{E}\Big[\exp\Big(p\sum_{i=1}^{n}\beta(t_{i-1})(W(t_i)-W(t_{i-1}))\Big)\Big]\\
&= \mathbb{E}\Big[\exp\Big(p\sum_{i=1}^{n-1}\beta(t_{i-1})(W(t_i)-W(t_{i-1}))\Big)\,\mathbb{E}[\exp(p\beta(t_{n-1})(W(t_n)-W(t_{n-1}))\,|\,\mathcal{F}_{t_{n-1}}]\Big]\\
&= \mathbb{E}\Big[\exp\Big(p\sum_{i=1}^{n-1}\beta(t_{i-1})(W(t_i)-W(t_{i-1}))\Big)\exp(p^2\beta(t_{n-1})^2(t_n-t_{n-1})\Big]\\
&\le \mathbb{E}\Big[\exp\Big(p\sum_{i=1}^{n-1}\beta(t_{i-1})(W(t_i)-W(t_{i-1}))\Big)\exp(p^2K^2(t_n-t_{n-1})\Big]\\
&\le \exp\Big(\sum_{i=1}^{n}p^2K^2(t_i-t_{i-1})\Big)\\
&= \exp(p^2K^2t).
\end{aligned}
$$

This shows that the random variables $\exp\Big(\sum_{i=1}^{n}\beta(t_{i-1})(W(t_i)-W(t_{i-1}))\Big)$ are uniformly bounded in any $L^p(\mathbb{P})$-space and thus uniformly integrable. Since by taking finer partitions these random variables converge to $\exp(\int_0^t\beta(s)\,dW(s))$ in $\mathbb{P}$-probability, we infer that $M(t)$ has finite expectation and even moments of all orders. Consequently, $\int_0^T\mathbb{E}[(\beta(t)M(t))^2]\,dt$ is finite and M is a martingale.

For the sufficency of Novikov's and Kazamaki's condition we refer to (Liptser and Shiryaev 2001) and the references and examples (!) there. □

3.3.5 Theorem. *Let $(X(t), 0\le t\le T)$ be a stochastic (Itô) process on $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ satisfying*

$$X(t) = \int_0^t \beta(s)\,ds + W(t),\quad 0\le t\le T,$$

with a Brownian motion W and a process $\beta\mathbf{1}_{t\le T} \in V^*$. If β is such that M is a martingale, then $(X(t), 0\le t\le T)$ is a Brownian motion under the measure $\mathbb{Q}$ on $(\Omega,\mathcal{F},(\mathcal{F}_t))$ defined by $\mathbb{Q}(d\omega) = M(T,\omega)\,\mathbb{P}(d\omega)$.

Proof. We use Lévy's characterisation of Brownian motion from Theorem 1.2.25. Since M is a martingale, $M(T)$ is a density and $\mathbb{Q}$ is well-defined.

We put $Z(t) = M(t)X(t)$ and obtain by Itô's formula (or partial integration)

$$\begin{aligned} dZ(t) &= M(t)\,dX(t) + X(t)\,dM(t) + d\langle M, X\rangle_t \\ &= M(t)\Big(\beta(t)\,dt + dW(t) - X(t)\beta(t)\,dW(t) - \beta(t)dt\Big) \\ &= M(t)(1 - X(t)\beta(t))\,dW(t). \end{aligned}$$

This shows that Z is a local martingale. If Z is a martingale, then we accomplish the proof using the preceding lemma:

$$\mathbb{E}_{\mathbb{Q}}[X(t)\,|\,\mathcal{F}_s] = \frac{\mathbb{E}_{\mathbb{P}}[M(t)X(t)\,|\,\mathcal{F}_s]}{\mathbb{E}_{\mathbb{P}}[M(t)\,|\,\mathcal{F}_s]} = \frac{Z(s)}{M(s)} = X(s), \quad s \le t,$$

implies that X is a $\mathbb{Q}$-martingale which by its very definition has quadratic variation t. Hence, X is a Brownian motion under $\mathbb{Q}$.

If Z is only a local martingale with associated stopping times (τ_n), then the above relation holds for the stopped processes $X^{\tau_n}(t) = X(t \wedge \tau_n)$, which shows that X is a local $\mathbb{Q}$-martingale and Lévy's theorem applies. □

3.3.6 Proposition. *Suppose X is a stochastic process on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ satisfying for some $T > 0$ and measurable functions b and σ*

$$dX(t) = b(X(t), t)\,dt + \sigma(X(t), t)\,dW(t), \quad 0 \le t \le T, \qquad X(0) = X_0.$$

Assume further that $u(x,t) := -c(x,t)/\sigma(x,t)$, c measurable, is such that

$$M(t) = \exp\Big(-\int_0^t u(X(s), s)\,dW(s) - \frac{1}{2}\int_0^t u^2(X(s), s)\,ds\Big), \quad 0 \le t \le T,$$

is an $(\mathcal{F}_t)$-martingale.

Then the stochastic differential equation

$$dY(t) = (b(Y(t), t) + c(Y(t), t))\,dt + \sigma(Y(t), t)\,dW(t), \quad 0 \le t \le T, \qquad Y(0) = X_0, \tag{3.3.1}$$

has a weak solution given by $((X, \widehat{W}), (\Omega, \mathcal{F}, \mathbb{Q}), (\mathcal{F}_t))$ for the $\mathbb{Q}$-Brownian motion

$$\widehat{W}(t) := W(t) + \int_0^t u(X(s), s)\,ds, \quad t \ge 0,$$

and the probability $\mathbb{Q}$ given by $\mathbb{Q}(d\omega) := M(T, \omega)\,\mathbb{P}(d\omega)$.

3.3.7 Remark. *Usually, the martingale $(M(t), t \ge 0)$ is not closable whence we are led to consider stochastic differential equations for finite time intervals.*

The martingale condition is for instance satisfied if σ is bounded away from zero and c is uniformly bounded. Putting $\sigma(x,t) = 1$ and $b(x,t) = 0$ we have weak existence for the equation $dX(t) = c(X(t), t)\,dt + dW(t)$ if c is Borel-measurable and satisfies a linear growth condition in the space variable, but without continuity assumption (Karatzas and Shreve 1991, Prop. 5.36).

Proof. From Theorem 3.3.5 we infer that $\widehat{W}$ is a $\mathbb{Q}$-Brownian motion. Hence, we can write

$$dX(t) = \Big(b(X(t),t) - \sigma(X(t),t)u(X(t),t)\Big)\,dt + \sigma(X(t),t)\,d\widehat{W}(t),$$

which by definition of u shows that $(X,\widehat{W})$ solves under $\mathbb{Q}$ equation (3.3.1). $\square$

The Girsanov Theorem also allows statements concerning uniqueness in law. The following is a typical version, which is proved in (Karatzas and Shreve 1991, Prop. 5.3.10, Cor 5.3.11).

3.3.8 Proposition. *Let two weak solutions $((X_i,W_i),(\Omega_i,\mathcal{F}_i,\mathbb{P}_i),(\mathcal{F}_t^i))$, $i=1,2$, of*

$$dX(t) = b(X(t),t)\,dt + dW(t), \quad 0 \le t \le T,$$

with $b:\mathbb{R}\times\mathbb{R}^+\to\mathbb{R}$ measurable be given with the same initial distribution. If $\mathbb{P}_i(\int_0^T |b(X_i(t),t)|^2\,dt < \infty) = 1$ holds for $i=1,2$, then (X_1,W_1) and (X_2,W_2) have the same law under the respective probability measures. In particular, if b is uniformly bounded, then uniqueness in distribution holds.

3.4 Applications in finance and statistics

Chapter 4

The Markov properties

4.1 General facts about Markov processes

Let us fix the measurable space (state space) $(S, \mathcal{S})$ and the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}; (\mathcal{F}_t)_{t\geq 0})$ until further notice. We present certain notions and results concerning Markov processes without proof and refer e.g. to Kallenberg (2002) for further information. We specialise immediately to processes in continuous time and later on also to processes with continuous trajectories.

4.1.1 Definition. *An S-valued stochastic process $(X(t),\, t \geq 0)$ is called Markov process if X is $(\mathcal{F}_t)$-adapted and satisfies*

$$\forall 0 \leq s \leq t,\ B \in \mathcal{S}:\ \mathbb{P}(X(t) \in B \,|\, \mathcal{F}_s) = \mathbb{P}(X(t) \in B \,|\, X(s)) \quad \mathbb{P}\text{-a.s.}$$

In the sequel we shall always suppose that regular conditional transition probabilities (Markov kernels) $\mu_{s,t}$ exist, that is for all $s \leq t$ the functions $\mu_{s,t} : S \times \mathcal{S} \to \mathbb{R}$ are measurable in the first component and probability measures in the second component and satisfy

$$\mu_{s,t}(X(s), B) = \mathbb{P}(X(t) \in B \,|\, X(s)) = \mathbb{P}(X(t) \in B \,|\, \mathcal{F}_s) \quad \mathbb{P}\text{-a.s.} \tag{4.1.1}$$

4.1.2 Lemma. *The Markov kernels $(\mu_{s,t})$ satisfy the Chapman-Kolmogorov equation*

$$\mu_{s,u}(x, B) = \int_S \mu_{t,u}(y, B)\, \mu_{s,t}(x, dy) \quad \forall\, 0 \leq s \leq t \leq u,\ x \in S,\ B \in \mathcal{S}.$$

4.1.3 Definition. *Any family of regular conditional probabilities $(\mu_{s,t})_{s\leq t}$ satisfying the Chapman-Kolmogorov equation is called a semigroup of Markov kernels. The kernels (or the associated process) are called time homogeneous if $\mu_{s,t} = \mu_{0,t-s}$ holds. In this case we just write μ_{t-s}.*

4.1.4 Theorem. *For any initial distribution ν auf $(S, \mathcal{S})$ and any semigroup of Markov kernels $(\mu_{s,t})$ there exists a Markov process X such that $X(0)$ is ν-distributed and equation (4.1.1) is satisfied.*

If S is a metric space with Borel σ-algebra $\mathcal{S}$ and if the process has a continuous version, then the process can be constructed on the path space $\Omega = C(\mathbb{R}^+, S)$ with its Borel σ-algebra $\mathfrak{B}$ and canonical right-continuous filtration $\mathcal{F}_t = \bigcap_{s>t} \sigma(X(u),\, u \leq s)$, where $X(u,\omega) := \omega(u)$ are the coordinate projections. The probability measure obtained is called $\mathbb{P}_\nu$ and it holds

$$\mathbb{P}_\nu = \int_S \mathbb{P}_x(A)\,\nu(dx), \quad A \in \mathfrak{B},$$

with $\mathbb{P}_x := \mathbb{P}_{\delta_x}$.

For the formal statement of the strong Markov property we introduce the shift operator ϑ_t that induces a left-shift on the function space Ω.

4.1.5 Definition. *The shift operator ϑ_t on the canonical space Ω is given by $\vartheta_t : \Omega \to \Omega$, $\vartheta_t(\omega) = \omega(t + \bullet)$ for all $t \geq 0$.*

4.1.6 Lemma.

1. *ϑ_t is measurable for all $t \geq 0$.*
2. *For $(\mathcal{F}_t)$-stopping times σ and τ the random time $\gamma := \sigma + \tau \circ \vartheta_\sigma$ is again an $(\mathcal{F}_t)$-stopping time.*

4.1.7 Theorem. *Let X be a time homogeneous Markov process and let τ be an $(\mathcal{F}_t)$-stopping time with at most countably many values. Then we have for all $x \in S$*

$$\mathbb{P}_x(X \circ \vartheta_\tau \in A \,|\, \mathcal{F}_\tau) = \mathbb{P}_{X(\tau)}(A) \quad \mathbb{P}_x\text{-a.s.} \quad \forall A \in \mathfrak{B}. \tag{4.1.2}$$

If X is the canonical process on the path space, then this is just an identity concerning the image measure under $\omega \mapsto \vartheta_{\tau(\omega)}(\omega)$: $\mathbb{P}_x(\bullet \,|\, \mathcal{F}_\tau) \circ (\vartheta_\tau)^{-1} = \mathbb{P}_{X(\tau)}$.

4.1.8 Definition. *A process X satisfying (4.1.2) for any finite (or equivalently bounded) stopping time τ is called strong Markov.*

4.1.9 Remark. *The strong Markov property entails the Markov property by setting $\tau = t$ and $A = \{X(s) \in B\}$ for some $B \in \mathcal{S}$ in (4.1.2).*

4.2 The martingale problem

We specify now to the state space $S = \mathbb{R}^d$. As before we work on the path space $\Omega = C(\mathbb{R}^+, \mathbb{R}^d)$ with its Borel σ-algebra $\mathfrak{B}$.

4.2.1 Definition. *A probability measure $\mathbb{P}$ on the path space $(\Omega, \mathfrak{B})$ is a solution of the local martingale problem for (b, σ) if*

$$M^f(t) := f(X(t)) - f(X(0)) - \int_0^t A_s f(X(s))\,ds, \quad t \geq 0,$$

where

$$A_s f(x) := \frac{1}{2}\sum_{i,j=1}^{d} (\sigma\sigma^T(x,s))_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \langle b(x,s), \operatorname{grad}(f)(x)\rangle,$$

$b : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^{d\times m}$ *measurable, is a local martingale under* $\mathbb{P}$ *for all functions* $f \in C_K^\infty(\mathbb{R}^d, \mathbb{R})$.

4.2.2 Remark. *If* b *and* σ *are bounded, then* $\mathbb{P}$ *even solves the martingale problem, for which* M^f *is required to be a proper martingale.*

4.2.3 Theorem. *The stochastic differential equation*

$$dX(t) = b(X(t),t)\,dt + \sigma(X(t),t)\,dW(t), \quad t \geq 0$$

has a weak solution $((X,W),(\Omega,\mathfrak{A},\mathbb{P}),(\mathfrak{F}_t))$ *if and only if a solution to the local martingale problem* (b,σ) *exists. In this case the law* $\mathbb{P}^X$ *of* X *on the path space equals the solution of the local martingale problem.*

Proof. For simplicity we only give the proof for the one-dimensional case, the multi-dimensional method of proof follows the same ideas.

1. Given a weak solution, Itô's rule yields for any $f \in C_K^\infty(\mathbb{R})$

   $$\begin{aligned} df(X(t)) &= f'(X(t))\,dX(t) + \tfrac{1}{2} f''(X(t))\,d\langle X\rangle_t \\ &= f'(X(t))\sigma(X(t),t)\,dW(t) + A_t f(X(t))\,dt. \end{aligned}$$

   Hence, M^f is a local martingale; just note that $\sigma(X(\bullet)) \in V^*$ is required for the weak solution and f' is bounded such that the stochastic integral is indeed well defined and a local martingale under $\mathbb{P}$. Of course, this remains true, when considered on the path space under the image measure $\mathbb{P}^X$.

2. Conversely, let $\mathbb{P}$ be a solution of the local martingale problem and consider functions $f_n \in C_K^\infty(\mathbb{R})$ with $f_n(x) = x$ for $|x| \leq n$. Then the standard stopping argument applied to M^{f_n} for $n \to \infty$ shows that

   $$M(t) := X(t) - X(0) - \int_0^t b(X(s),s)\,ds, \quad t \geq 0,$$

   is a local martingale. Similarly approximating $g(x) = x^2$, we obtain that

   $$N(t) := X^2(t) - X^2(0) - \int_0^t \sigma^2(X(s),s) + b(X(s),s)2X(s)\,ds, \quad t \geq 0,$$

   is a local martingale. By Itô's formula, $dX^2(t) = 2X(t)dX(t) + d\langle X\rangle_t$ holds and shows

   $$N(t) = \int_0^t 2X(s)\,dM(s) + \langle M\rangle_t - \int_0^t \sigma^2(X(s),s)\,ds, \quad t \geq 0.$$

Therefore $\langle M\rangle_t - \int_0^t \sigma^2(X(s),s)\,ds$ is a continuous local martingale of bounded variation. By (Revuz and Yor 1999, Prop. IV.1.2) it must therefore vanish identically and $d\langle M\rangle_t = \sigma^2(X(t),t)\,dt$ follows. By the representation theorem for continuous local martingales (Kallenberg 2002, Thm. 18.12) there exists a Brownian motion W such that $M(t) = \int_0^t \sigma(X(s),s)\,dW(s)$ holds for all $t \geq 0$. Consequently (X,W) solves the stochastic differential equation.

□

4.2.4 Corollary. *A stochastic differential equation has a (in distribution) unique weak solution if and only if the corresponding local martingale problem is uniquely solvable, given some initial distribution.*

4.3 The strong Markov property

We immediately start with the main result that solutions of stochastic differential equations are under mild conditions strong Markov processes. This entails that the solutions are diffusion processes in the sense of Feller (Feller 1971).

4.3.1 Theorem. *Let $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be time-homogeneous measurable coefficients such that the local martingale problem for (b,σ) has a unique solution $\mathbb{P}_x$ for all initial distributions δ_x, $x \in \mathbb{R}^d$. Then the family $(\mathbb{P}_x)$ satisfies the strong Markov property.*

Proof. In order to state the strong Markov property we need that $(\mathbb{P}_x)_{x\in\mathbb{R}^d}$ are Markov kernels. Theorem 21.10 of Kallenberg (2002) shows by abstract arguments that $x \mapsto \mathbb{P}_x(B)$ is measurable for all $B \in \mathfrak{B}$.

We thus have to show

$$\mathbb{P}_x(X \circ \vartheta_\tau \in B \,|\, \mathcal{F}_\tau) = \mathbb{P}_{X(\tau)}(B) \quad \mathbb{P}_x\text{-a.s. } \forall\, B \in \mathfrak{B},\ \text{bounded stopping time } \tau.$$

By the unique solvability of the martingale problem it suffices to show that the random (!) probability measure $\mathbb{Q}_\tau := \mathbb{P}_x((\vartheta_\tau)^{-1}\bullet \,|\, \mathcal{F}_\tau)$ solves $\mathbb{P}_x$-almost surely the martingale problem for (b,σ) with initial distribution $\delta_{X(\tau)}$. Concerning the initial distribution we find for any Borel set $A \subset \mathbb{R}^d$ by the stopping time property of τ

$$\begin{aligned}
\mathbb{P}_x((\vartheta_\tau)^{-1}\{\omega' \,|\, \omega'(0) \in A\} \,|\, \mathcal{F}_\tau)(\omega) &= \mathbb{P}_x(\{\omega' \,|\, \omega'(\tau(\omega')) \in A\} \,|\, \mathcal{F}_\tau)(\omega)\\
&= \mathbf{1}_A(\omega(\tau(\omega)))\\
&= \mathbf{1}_A(X(\tau(\omega),\omega))\\
&= \mathbb{P}_{X(\tau(\omega),\omega)}(\{\omega' \,|\, \omega'(0) \in A\}).
\end{aligned}$$

It remains to prove the local martingale property of M^f under $\mathbb{Q}_\tau$, that is the martingale property of $M^{f,n}(t) := M^f(t \wedge \tau_n)$ with $\tau_n := \inf\{t \geq 0 \,|\, \|M^f(t)\| \leq n\}$.

By its very definition $M^f(t)$ is always $\mathcal{F}_t$-measurable, so we prove that $\mathbb{P}_x$-almost surely

$$\int_F M^{f,n}(t,\omega')\,\mathbb{Q}_\tau(d\omega') = \int_F M^{f,n}(s,\omega')\,\mathbb{Q}_\tau(d\omega') \quad \forall F \in \mathcal{F}_s,\ s \le t.$$

By the separability of Ω and the continuity of $M^{f,n}$ it suffices to prove this identity for countably many F, s and t (Kallenberg 2002, Thm. 21.11). Consequently, we need not worry about $\mathbb{P}_x$-null sets. We obtain

$$\begin{aligned}\int_F M^{f,n}(t,\omega')\,\mathbb{Q}_\tau(d\omega') &= \int \mathbf{1}_F(\vartheta_\tau(\omega''))M^{f,n}(t,\vartheta_\tau(\omega''))\ \mathbb{P}_x(d\omega''\,|\,\mathcal{F}_\tau)\\ &= \mathbb{E}_x[\mathbf{1}_{(\vartheta_\tau)^{-1}F}M^{f,n}(t,\vartheta_\tau)\,|\,\mathcal{F}_\tau].\end{aligned}$$

Because of $M^{f,n}(t,\vartheta_\tau) = M^f((t+\tau)\wedge\sigma_n)$ with $\sigma_n := \tau_n\circ\vartheta_\tau + \tau$, which is by Lemma 4.1.6 a stopping time, the process $M^{f,n}(t,\vartheta_\tau)$ is a martingale under $\mathbb{P}_x$ adapted to $(\mathcal{F}_{t+\tau})_{t\ge 0}$. Since $(\vartheta_\tau)^{-1}F$ is an element of $\mathcal{F}_{s+\tau}$, we conclude by optional stopping that $\mathbb{P}_x$-almost surely

$$\begin{aligned}\int_F M^{f,n}(t,\omega')\,\mathbb{Q}_\tau(d\omega') &= \mathbb{E}_x[\mathbf{1}_{(\vartheta_\tau)^{-1}F}\,\mathbb{E}_x[M^{f,n}(t,\vartheta_\tau)\,|\,\mathcal{F}_{s+\tau}]\,|\,\mathcal{F}_\tau]\\ &= \mathbb{E}_x[\mathbf{1}_{(\vartheta_\tau)^{-1}F}M^{f,n}(s+\tau)\,|\,\mathcal{F}_\tau]\\ &= \int_F M^{f,n}(s,\omega')\,\mathbb{Q}_\tau(d\omega').\end{aligned}$$

Consequently, we have shown that with $\mathbb{P}_x$-probability one $\mathbb{Q}_\tau$ solves the martingale problem with initial distribution $X(\tau)$ and therefore equals $\mathbb{P}_{X(\tau)}$. □

4.3.2 Example. *A famous application is the reflection principle for Brownian motion W. By the strong Markov property, for any finite stopping time τ the process $(W(t+\tau)-W(\tau),\ t\ge 0)$ is again a Brownian motion independent of $\mathcal{F}_\tau$ such that with $\tau_b := \inf\{t\ge 0\,|\,W(t)\ge b\}$ for some $b>0$:*

$$\begin{aligned}\mathbb{P}_0(\tau_b\le t) &= \mathbb{P}_0(\tau_b\le t,\ W(t)\ge b) + \mathbb{P}_0(\tau_b\le t,\ W(t)<b)\\ &= \mathbb{P}_0(W(t)\ge b) + \mathbb{P}_0(\tau_b\le t,\ W(\tau_b+(t-\tau_b)) - W(\tau_b) < 0)\\ &= \mathbb{P}_0(W(t)\ge b) + \tfrac{1}{2}\,\mathbb{P}_0(\tau_b\le t).\end{aligned}$$

This implies $\mathbb{P}_0(\tau_b\le t) = 2\,\mathbb{P}(W(t)>b)$ and the stopping time τ_b has a distribution with density

$$f_b(t) = \frac{b}{\sqrt{2\pi t^3}}e^{-b^2/(2t)}, \quad t\ge 0.$$

Because of $\{\tau_b\le t\} = \{\max_{0\le s\le t} W(t)\ge b\}$ we have at the same time determined the distribution of the maximum of Brownian motion on any finite interval.

4.4 The infinitesimal generator

We first gather some facts concerning Markov transition operators and their semigroup property, see Kallenberg (2002) or Revuz and Yor (1999).

4.4.1 Lemma. *Given a family $(\mu_t)_{t\ge 0}$ of time-homogeneous Markov kernels, the operators*

$$T_t f(x) := \int f(y)\mu_t(x,dy), \ \ f : S \to \mathbb{R} \text{ bounded, measurable,}$$

form a semigroup, that is $T_t \circ T_s = T_{t+s}$ holds for all $t, s \ge 0$.

Proof. Use the Chapman-Kolmogorov equation. □

We now specialise to the state space $S = \mathbb{R}^d$ with its Borel σ-algebra.

4.4.2 Definition. *If the operators $(T_t)_{t\ge 0}$ satisfy (a) $T_t f \in C_0(\mathbb{R}^d)$ for all $f \in C_0(\mathbb{R}^d)$ and (b) $\lim_{h\to 0} T_h f(x) = f(x)$ for all $f \in C_0(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, then (T_t) is called a Feller semigroup.*

4.4.3 Theorem. *A Feller semigroup $(T_t)_{t\ge 0}$ is a strongly continuous operator semigroup on $C_0(\mathbb{R}^d)$, that is $\lim_{h\to 0} T_h f = f$ holds in supremum norm. It is uniquely determined by its generator $A : \mathcal{D}(A) \subset C_0(\mathbb{R}^d) \to \mathbb{R}^d$ with*

$$Af := \lim_{h\to 0} \frac{T_h f - f}{h}, \quad \mathcal{D}(A) := \{f \in C_0(\mathbb{R}^d) \mid \lim_{h\to 0} \tfrac{T_h f - f}{h} \text{ exists}\}.$$

Moreover, the semigroup uniquely defines the Markov kernels and thus the distribution of the associated Markov process (which is called Feller process).

4.4.4 Corollary. *We have for all $f \in \mathcal{D}(A)$*

$$\tfrac{d}{dt} T_t f = AT_t f = T_t Af.$$

4.4.5 Theorem. *(Hille-Yosida) Let A be a closed linear operator on $C_0(\mathbb{R}^d)$ with dense domain $\mathcal{D}(A)$. Then A is the generator of a Feller semigroup if and only if*

1. *the range of $\lambda_0 \operatorname{Id} - A$ is dense in $C_0(\mathbb{R}^d)$ for some $\lambda_0 > 0$;*
2. *if for some $x \in \mathbb{R}^d$ and $f \in \mathcal{D}(A)$, $f(x) \ge 0$ and $f(x) = \max_{y\in\mathbb{R}^d} f(x)$ then $Af(x) \le 0$ follows (positive Maximum principle).*

4.4.6 Theorem. *If b and σ are bounded and satisfy the conditions of Theorem 4.3.1, then the Markov kernels $(\mathbb{P}_x)_{x\in\mathbb{R}^d}$ solving the martingale problem for (b,σ) give rise to a Feller semigroup (T_t). Any function $f \in C_0^2(\mathbb{R}^d)$ lies in $\mathcal{D}(A)$ and fulfills*

$$Af(x) = \frac{1}{2}\sum_{i,j=1}^{d} (\sigma\sigma^T(x))_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \langle b(x), \operatorname{grad}(f)(x)\rangle.$$

We shall even prove a stronger result under less restrictive conditions, which turns out to be a very powerful tool in calculating certain distributions for the solution processes.

4.4.7 Theorem. *(Dynkin's formula) Assume that b and σ are measurable, locally bounded and such that the SDE* (2.1.1) *with time-homogeneous coefficients has a (in distribution) unique weak solution. Then for all $x \in \mathbb{R}^d$, $f \in C_K^2(\mathbb{R}^d)$ and all bounded stopping times τ we have*

$$\mathbb{E}_x[f(X(\tau))] = f(x) + \mathbb{E}_x\Big[\int_0^\tau Af(X(s))\,ds\Big].$$

Proof. By Theorem 4.2.3 the process M^f is a local martingale under $\mathbb{P}_x$. By the compact support of f and the local boundedness of b and σ we infer that $M^f(t)$ is uniformly bounded and therefore M^f is a martingale. Then the optional stopping result $\mathbb{E}[M^f(\tau)] = \mathbb{E}[M^f(0)] = 0$ yields Dynkin's formula. □

4.4.8 Example.

1. *Let W be an m-dimensional Brownian motion starting in some point a and $\tau_R := \inf\{t \geq 0 \,|\, \|W(t)\| \geq R\}$. Then $\mathbb{E}_a[\tau_R] = (R^2 - \|a\|^2)/m$ holds for $\|a\| < R$. To infer this from Dynkin's formula put $f(x) = \|x\|^2$ for $\|x\| \leq R$ and extend f outside of the ball such that $f \in C^2(\mathbb{R})$ with compact support. Then $Af(x) = m$ for $\|x\| \leq R$ and therefore Dynkin's formula yields $\mathbb{E}_a[f(W(\tau_R \wedge n))] = f(a) + m\,\mathbb{E}_a[\tau_R \wedge n]$. By montone convergence,*

$$\mathbb{E}_a[\tau_R] = \lim_{n\to\infty} \mathbb{E}_a[\tau_R \wedge n] = \lim_{n\to\infty} (\mathbb{E}_a[\|W(\tau_R \wedge n)\|^2] - \|a\|^2)/m$$

   holds and we can conclude by dominated convergence ($\|W(\tau_R \wedge n)\| \leq R$).

2. *Consider the one-dimensional stochastic differential equation*

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

   Suppose a weak solution exists for some initial value $X(0)$ with $\mathbb{E}[X(0)^2] < \infty$ and that $\sigma^2(x) + 2xb(x) \leq C(1+x^2)$ holds. Then $\mathbb{E}[X(t)^2] \leq (\mathbb{E}[X(0)^2]+1)e^{Ct} - 1$ follows. To prove this, use the same f and put $\kappa_t := \tau_R \wedge t$ with τ_R from above for all $t \geq 0$ such that by Dynkin's formula

$$\mathbb{E}_x[X(\kappa_t)^2] = x^2 + \mathbb{E}_x\Big[\int_0^{\kappa_t} \big(\sigma^2(X(s)) + 2b(X(s))X(s)\big)\,ds\Big] \leq x^2 + \int_0^t C(1 + X(s\wedge\kappa)^2)\,ds.$$

   By Gronwall' s lemma, we obtain $\mathbb{E}_x[1 + X(\kappa_t)^2] \leq (x^2+1)e^{Ct}$. Since this is valid for any $R > 0$ we get $\mathbb{E}_x[X(t)^2] \leq (x^2+1)e^{Ct} - 1$ and averaging over the initial condition yields $\mathbb{E}[X(t)^2] \leq (\mathbb{E}[X(0)^2]+1)e^{Ct} - 1$. Note that this kind of approach was already used in Theorem 2.3.3 and improves significantly on the moment estimate of Theorem 2.3.1.

3. *For the solution process X of a one-dimensional SDE as before we consider the stopping time $\tau := \inf\{t \geq 0 \,|\, X(t) = 0\}$. We want to decide whether $\mathbb{E}_a[\tau]$ is finite or infinite for $a > 0$. For this set $\tau_R := \tau \wedge \inf\{t \geq 0 \,|\, X(t) \geq R\}$, $R > a$, and consider a function $f \in C^2(\mathbb{R})$ with compact support, $f(0) = 0$ and solving $Af(x) = 1$ for $x \in [0, R]$. Then Dynkin's formula yields*
$$\mathbb{E}_a[f(X(\tau_R \wedge n))] = f(a) + \mathbb{E}_a[\tau_R \wedge n].$$
For a similar function g with $Ag = 0$ and $g(0) = 0$ we obtain $\mathbb{E}_a[g(X(\tau_R \wedge n))] = g(a)$. Hence,
$$\begin{aligned}\mathbb{E}_a[\tau_R \wedge n] &= \mathbb{E}_a[f(X(\tau_R \wedge n))] - f(a)\\ &= \mathbb{P}_a(X(\tau_R \wedge n) = R) f(R) + \mathbb{E}_a[f(X(n))\mathbf{1}_{\{\tau_R > n\}}] - f(a)\\ &= \Big(g(a) - \mathbb{E}_a[g(X(n))\mathbf{1}_{\{\tau_R > n\}}]\Big)\frac{f(R)}{g(R)} + \mathbb{E}_a[f(X(n))\mathbf{1}_{\{\tau_R > n\}}] - f(a)\end{aligned}$$
follows. Using the uniform boundedness of f and g we infer by montone and dominated convergence for $n \to \infty$
$$\mathbb{E}_a[\tau_R] = g(a)\frac{f(R)}{g(R)} - f(a).$$
Monotone convergence for $R \to \infty$ thus gives $\mathbb{E}_a[\tau] < \infty$ if and only if $\lim_{R\to\infty} \frac{f(R)}{g(R)}$ is finite. The functions f and g can be determined in full generality, but we restrict ourselves to the case of vanishing drift $b(x) = 0$ and strictly positive diffusion coefficient $\inf_{0 \leq y \leq x} \sigma(y) > 0$ for all $x > 0$. Then
$$f(x) = \int_0^x \int_0^y \frac{2}{\sigma^2(z)}\, dz\, dy \text{ and } g(x) = x$$
will do. Since $f(x) \to \infty$, $g(x) \to \infty$ hold for $x \to \infty$, L'Hopital's rule gives
$$\lim_{R\to\infty} \frac{f(R)}{g(R)} = \lim_{R\to\infty} \frac{f'(R)}{g'(R)} = \int_0^\infty \frac{2}{\sigma^2(z)}\, dz.$$
We conclude that the solution of $dX(t) = \sigma(X(t))\, dW(t)$ with $X(0) = a$ satisfies $\mathbb{E}_a[\tau] < \infty$ if and only if σ^{-2} is integrable. For constant σ we obtain a multiple of Brownian motion which satisfies $\mathbb{E}_a[\tau] = \infty$. For $\sigma(x) = x + \varepsilon$, $\varepsilon > 0$, $\mathbb{E}_a[\tau] < \infty$ holds, but in the limit $\varepsilon \to 0$ the expectation tends to infinity. This can be understood when observing that a solution of $dX(t) = (X(t) + \varepsilon) dW(t)$ is given by the translated geometric Brownian motion $X(t) = \exp(W(t) - \frac{t}{2}) - \varepsilon$, which tends to $-\varepsilon$ almost surely, but never reaches the value $-\varepsilon$. Concerning the behaviour of $\sigma(x)$ for $x \to \infty$ we note that $\mathbb{E}_a[\tau]$ is finite as soon as $\sigma(x)$ grows at least like x^α for some $\alpha > \frac{1}{2}$ such that the rapid fluctuations of X for large x make excursions towards zero more likely.

4.5 The Kolmogorov equations

The main object one is usually interested in to calculate for the solution process X of an SDE is the transition probability $\mathbb{P}(X(t) \in B \,|\, X(s) = x)$ for $t \geq s \geq 0$ and any Borel set B. A concise description is possible, if a transition density $p(x, y; t)$ exists satisfying

$$\mathbb{P}(X(t) \in B \,|\, X(s) = x) = \int_B p(x, y; t - s)\, dy.$$

Here we shall present analytical tools to determine this transition density if it exists. The proof of its existence usually either relies completely on analytical results or on Malliavin calculus, both being beyond our scope.

4.5.1 Lemma. *Assume that b and σ are continuous and such that the SDE* (2.1.1) *has a (in distribution) unique weak solution for any deterministic initial value. For any $f \in C_K^2(\mathbb{R}^d)$ set $u(x,t) := \mathbb{E}_x[f(X(t))]$. Then u is a solution of the parabolic partial differential equation*

$$\frac{\partial u}{\partial t}(x,t) = (Au(\bullet, t))(x), \quad \forall\, x \in \mathbb{R}^d, t \geq 0, \text{ with } u(x,0) = f(x) \quad \forall\, x \in \mathbb{R}^d.$$

Proof. Dynkin's formula for $\tau = t$ yields by the Fubini-Tonelli theorem

$$u(x,t) = f(x) + \int_0^t \mathbb{E}_x[Af(X(s))]\, ds \quad \forall\, x \in \mathbb{R}^d, t \geq 0.$$

Since the coefficients b and σ are continuous, the integrand is continuous and u is continuously differentiable with respect to t satisfying $\frac{\partial u}{\partial t}(x,t) = \mathbb{E}_x[Af(X(t))]$. On the other hand we obtain by the Markov property for $t, h > 0$

$$\mathbb{E}_x[u(X(h), t)] = \mathbb{E}_x[\mathbb{E}_{X(h)}[f(X(t))]] = \mathbb{E}_x[f(X(t+h))] = u(x, t+h).$$

For fixed $t > 0$ we infer that the left hand side of

$$\frac{u(x, t+h) - u(x,t)}{h} = \frac{\mathbb{E}_x[u(X(h), t)] - u(x,t)}{h}$$

converges for $h \to 0$ to $\frac{\partial u}{\partial t}$ and therefore also the right-hand side. Therefore u lies in the domain $\mathcal{D}(A)$ and the assertion follows. □

4.5.2 Corollary. *If the transition density $p(x, y; t)$ exists, is twice continuously differentiable with respect to x and continuously differentiable with respect to t, then $p(x, y; t)$ solves for all $y \in \mathbb{R}^d$ the backward Kolmogorov equation*

$$\frac{\partial u}{\partial t}(x,t) = (Au(\bullet, t))(x), \quad \forall\, x \in \mathbb{R}^d, t \geq 0, \text{ with } u(x,0) = \delta_y(x).$$

In other words, for fixed y the transition density is the fundamental solution of this parabolic partial differential equation.

Proof. Writing the identity in the preceding lemma in terms of p, we obtain for any $f \in C_K^2(\mathbb{R}^d)$

$$\frac{\partial}{\partial t}\int_{\mathbb{R}^d} f(y)p(x,y;t)\,dy = A\Big(\int_{\mathbb{R}^d} f(y)p(x,y;t)\,dy\Big).$$

By the compact support of f and the smoothness properties of p, we may interchange integration and differentiation on both sides. From $\int(\frac{\partial}{\partial t} - A)p(x,y;t)f(y)dy = 0$ for any test function f we then conclude by a continuity argument. □

4.5.3 Corollary. *If the transition density $p(x,y;t)$ exists, is twice continuously differentiable with respect to y and continuously differentiable with respect to t, then $p(x,y;t)$ solves for all $x \in \mathbb{R}^d$ the forward Kolmogorov equation*

$$\frac{\partial u}{\partial t}(y,t) = (A^* u(\bullet,t))(y), \quad \forall\, y \in \mathbb{R}^d, t \geq 0, \text{ with } u(y,0) = \delta_x(y),$$

where

$$A^* f(y) = \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial y^2}\Big((\sigma\sigma^T(y))_{ij} f(y)\Big) - \sum_{i=1}^d \frac{\partial}{\partial y_i}\Big(b_i(y)f(y)\Big)$$

is the formal adjoint of A. Hence, for fixed x the transition density is the fundamental solution of the parabolic partial differential equation with the adjoint operator.

Proof. Let us evaluate $\mathbb{E}_x[Af(X(t))]$ for any $f \in C_K^2(\mathbb{R}^d)$ in two different ways. First, we obtain by definition

$$\mathbb{E}_x[Af(X(t))] = \int Af(y)p(x,y;t)\,dy = \int f(y)(A^* p(x,\bullet;t))(y)\,dy.$$

On the other hand, by dominated convergence and by Dynkin's formula we find

$$\int f(y)\frac{\partial}{\partial t}p(x,y;t)\,dy = \frac{\partial}{\partial t}\,\mathbb{E}_x[f(X(t))] = \mathbb{E}_x[Af(X(t))].$$

We conclude again by testing this identity with all $f \in C_K^2(\mathbb{R}^d)$. □

4.5.4 Remark. *The preceding results are in a sense not very satisfactory because we had to postulate properties of the unknown transition density in order to derive a determining equation. Karatzas and Shreve (1991) state on page 368 sufficient conditions on the coefficients b and σ, obtained from the analysis of the partial differential equations, under which the transition density is the unique classical solution of the forward and backward Kolmogorov equation, respectively. Main hypotheses are ellipticity of the diffusion coefficient and boundedness of both coefficients together with certain Hölder-continuity requirements. In the case of the forward equation in addition the first two derivatives of σ and the first derivative of b have to have these properties, which is intuitively explained by the form of the adjoint A^*.*

4.5.5 Example. *We have seen that a solution of the scalar Ornstein-Uhlenbeck process*

$$dX(t) = \alpha X(t)\,dt + \sigma\,dW(t), \quad t \geq 0,$$

is given by $X(t) = X(0)e^{\alpha t} + \int_0^t e^{\alpha(t-s)}\sigma\ dW(s)$. *Hence, the transition density is given by the normal density*

$$p(x,y;t) = \frac{1}{\sqrt{2\pi\sigma^2(2\alpha)^{-1}(e^{2\alpha t}-1)}} \exp\Big(-(y-xe^{\alpha t})^2/(\sigma^2\alpha^{-1}(e^{2\alpha t}-1))\Big).$$

It can be easily checked that p *solves the Kolmogorov equations*

$$\frac{\partial u}{\partial t}(x,t) = \frac{\sigma^2}{2}\frac{\partial^2 u}{\partial x^2}(x,t) + \alpha\frac{\partial u}{\partial x}(x,t) \qquad \text{and} \qquad \frac{\partial u}{\partial t}(y,t) = \frac{\sigma^2}{2}\frac{\partial^2 u}{\partial x^2}(y,t) - \alpha\frac{\partial u}{\partial y}(y,t).$$

For $\alpha = 0$ *and* $\sigma = 1$ *we obtain the Brownian motion transition density* $p(x,y;t) = (2\pi t)^{-1/2}\exp((y-x)^2/(2t))$ *which is the fundamental solution of the classical heat equation* $\frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}$ *in both variables* x *and* y.

4.6 The Feynman-Kac formula

Chapter 5

Stochastic control: an outlook

In this chapter we briefly present one main approach for solving optimal control problems for dynamical systems described by stochastic differential equations: Bellman's principle of dynamic programming and the resulting Hamilton-Jacobi-Bellman equation.

For some $T > s \geq 0$ and $y \in \mathbb{R}^d$ we consider the controlled stochastic differential equation

$$dX(t) = b(X(t), u(t), t)\,dt + \sigma(X(t), u(t), t)\,dW(t), \quad t \in [s, T], \quad X(s) = y,$$

where X is d-dimensional, W is m-dimensional Brownian motion and the coefficients $b : \mathbb{R}^d \times U \times [0, T] \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times U \times [0, T] \to \mathbb{R}^{d\times m}$ are regular, say Lipschitz continuous, in x and depend on the controls $u(t)$ taken in some abstract metric space U, which are $\mathcal{F}_t$-adapted. The goal is to choose the control u in such a way that a given cost functional

$$J(s, y; u) := \mathbb{E}\Big[\int_s^T f(X(t), u(t), t)\,dt + h(X(t))\Big],$$

where f and h are certain continuous functions, is minimized.

5.0.1 Example. *A standard example of stochastic control is to select a portfolio of assets, which is in some sense optimal. Suppose a riskless asset S_0 like a bond grows by a constant rate $r > 0$ over time*

$$dS_0(t) = rS_0(t)\,dt,$$

while a risky asset S_1 like a stock follows the scalar diffusion equation of a geometric Brownian motion (Black-Scholes model)

$$dS_1(t) = S_1(t)\Big(b\,dt + \sigma\,dW(t)\Big).$$

Since this second asset is risky, it is natural to suppose $b > r$. The agent has at each time t the possibility to trade, that is to decide the fraction $u(t)$ of his wealth $X(t)$

which is invested in the risky asset S_1. Under this model we can derive the stochastic differential equation governing the dynamics of the agent's wealth:

$$\begin{aligned} dX(t) &= u(t)X(t)\Big(b\,dt + \sigma\,dW(t)\Big) + (1-u(t))X(t)r\,dt \\ &= (r + (b-r)u(t))X(t)\,dt + \sigma u(t)X(t)\,dW(t). \end{aligned}$$

Note that necessarily $u(t) \in [0,1]$ has to hold for all t and we should choose $U = [0,1]$. Suppose the investor wants to maximize his average utility at time $T > 0$, where the utility is usually assumed to be a concave function of the wealth. Then a mathematically tractable cost functional would for instance be

$$J(s,y;u) = -\,\mathbb{E}_{s,y;u}[X(T)^\alpha], \qquad \alpha \in (0,1].$$

Note that the expectation depends of course on the initial wealth endowment $X(s) = y$ and the chosen investment strategy u. The special linear form of the system implies automatically that the wealth process X cannot become negative and the cost funtional is well-defined., but usually one has to treat this kind of restriction separately, either by specifying the set of admissible controls more precisely or by introducing a stopping time instead of the deterministic final time T.

We will not write down properly all assumptions on b, σ, f and h, but refer to (Yong and Zhou 1999, Section 4.3.1) for details, in particular the discussion about the requirement of having a weak or a strong solution of the SDE involved. Here, we only want to stress the fact that we allow all controls u in a set of admissible controls $U(s,T) \subset \{u : [s,T] \times \Omega \to U \,|\, u(t,\bullet) \text{ is } \mathcal{F}_t^s \text{ adapted}\}$, where $\mathcal{F}_t^s$ is generated by the Brownian motion W in the period $[s,t]$ and augmented by null sets.

The problem of optimal stochastic control can then be stated as follows: Find for given (s,y) a control process $\bar{u} \in U(s,T)$ such that

$$J(s,y;\bar{u}) = \inf_{u \in U(s,T)} J(s,y;u).$$

The existence of an optimal control process is not always ensured, but in many cases follows from the setup of the problem or by compactness arguments.

We can now state the main tool we want to use for solving this optimization problem. Conceptually, the idea is to study how the optimal cost changes over time and state. This means that we shall consider the so-called value function

$$V(s,y) := \inf_{u \in U(s,T)} J(s,y;u), \quad (s,y) \in [0,T) \times \mathbb{R}^d,$$

with its natural extension $V(T,y) = h(y)$.

5.0.2 Theorem. *(Bellman's dynamic programming principle) Under certain regularity conditions we have for any $(s,y) \in [0,T) \times \mathbb{R}^d$ and $z \in [s,T]$:*

$$V(s,y) = \inf_{u \in U(s,T)} \mathbb{E}\Big[\int_s^z f(X^{s,y,u}(t), u(t), t)\,dt + V(z, X^{s,y,u}(z))\Big].$$

Proof. See Theorem 4.3.3 in Yong and Zhou (1999). □

Intuitively, this principle asserts that a globally optimal control u over the period $[s,T]$ is also locally optimal for shorter periods $[u,T]$. In other words, we cannot improve upon a globally optimal control by optimising separately on smaller subintervals. If this were the case, we could simply patch together these controls to obtain a globally better control.

The key point is that the knowledge of the value function for all arguments allows to determine also the optimal controls which have to be applied in order to attain the optimal cost. Therefore we have to study the equation for V in the Bellman principle more thoroughly. Since integral equations are more difficult to handle, we look for infinitesimal changes in s, which amounts to letting $z \downarrow s$ appropriately. Heuristically, we interchange limit and infimum in the following formal(!) calculations, which have to be justified much more accurately:

$$0 = \frac{1}{z-s} \inf_{u \in U(s,T)} \mathbb{E}\Big[\int_s^z f(X^{s,y,u}(t), u(t), t)\,dt + V(z, X^{s,y,u}(z)) - V(s,y)\Big]$$

then gives formally for $z \downarrow s$

$$0 = \inf_{u \in U(s,T)} \Big(f(y, u(s), s) + \frac{\partial}{\partial t} \mathbb{E}[V(t, X^{s,y,u}(t))]|_{t=s} \Big),$$

which using the theory developed in the preceding chapter yields

$$0 = \inf_{u \in U} \Big(f(y,u,s) + \frac{\partial}{\partial s} V(s,y) + A^{s,y,u} V(s,y) \Big),$$

where we have denoted by $A^{s,y,u}$ the infinitesimal generator associated to $X^{s,y,u}$:

$$A^{s,y,u} f(y) = \frac{1}{2} \sum_{i,j=1}^d (\sigma\sigma^T(s,y,u))_{ij} \frac{\partial^2 f}{\partial y_i \partial y_j}(y) + \langle b(s,y,u), \operatorname{grad}(f)(y) \rangle.$$

In terms of the so-called Hamiltonian

$$H(t,x,u,p,P) := \frac{1}{2} \operatorname{trace}(P(\sigma\sigma^T)(s,y,u)) + \langle b(s,y,u), p \rangle - f(s,y,u)$$

we arrive at the Hamilton-Jacobi-Bellman (HJB) equation

$$\frac{\partial V}{\partial s} = \sup_{u \in U} H\big(s, y, u, -\big(\tfrac{\partial V}{\partial y_i}\big)_i, -\big(\tfrac{\partial^2 V}{\partial y_i \partial y_j}\big)_{ij}\big), \quad (s,y) \in [0,T) \times \mathbb{R}^d.$$

Together with $V(T,y) = h(y)$ we thus focus a terminal value problem for a partial differential equation.

In general, the value function only solves the HJB equation in a weak sense as a so called viscosity solution.

In the sequel we assume that we have found the value function, e.g. via solving the HJB equation and proving uniqueness of the solution in a certain sense. Then the optimal control $\bar{u}$ is given in feedback form $\bar{u}(t) = u^*(X(t), t)$ with u^* found by the maximizing property

$$H\big(s, y, u^*(y,s), -\big(\tfrac{\partial V}{\partial y_i}\big)_i, -\big(\tfrac{\partial^2 V}{\partial y_i \partial y_j}\big)_{ij}\big) = sup_{u\in U} H\big(s, y, u, -\big(\tfrac{\partial V}{\partial y_i}\big)_i, -\big(\tfrac{\partial^2 V}{\partial y_i \partial y_j}\big)_{ij}\big).$$

For a correct mathematical statement we cite the standard classical verification theorem from (Yong and Zhou 1999, Thm. 5.5.1).

5.0.3 Theorem. *Suppose $W \in C^{1,2}([0,T], \mathbb{R}^d)$ solves the HJB equation together with its final value. Then*

$$W(s,y) \le J(s,y;u)$$

holds for all controls u and all (s,y), that is W is a lower bound for the value function. Furthermore, an admissible control $\bar{u}$ is optimal if and only if

$$\frac{\partial V}{\partial t}(t, X^{s,y,\bar{u}}(t)) = H\big(t, X^{s,y,\bar{u}}(t), \bar{u}(t), -\big(\tfrac{\partial V}{\partial y_i}(t, X^{s,y,\bar{u}}(t))\big)_i, -\big(\tfrac{\partial^2 V}{\partial y_i \partial y_j}(t, X^{s,y,\bar{u}}(t))\big)_{ij}\big)$$

holds for $t \in [s,T]$ almost surely.

Let us close this chapter by reconsidering the optimal investment example. The Hamiltonian in this case is given by

$$H(t,x,u,p,P) = \frac{1}{2}\sigma^2 u^2 x^2 P + (r + (b-r)u)xp$$

such that the HJB equation reads

$$\partial_t V(t,x) = \sup_{u\in[0,1]} \Big(-\frac{1}{2}\sigma^2 u^2 x^2 \partial_{xx} V(t,x) - (r + (b-r)u)x\partial_x V(t,x)\Big).$$

Neglecting for a moment the restriction $u \in [0,1]$ we find the optimizing value u^* in this equation by the first order condition

$$\sigma^2 u^* x^2 \partial_{xx} V(t,x) + (b-r)x\partial_x V(t,x) = 0$$

leading to the more explicit HJB equation

$$\begin{aligned}\partial_t V(t,x) &= -\frac{1}{2}\frac{(r-b)^2 x^2 (\partial_x V)^2}{\sigma^2 x^2 \partial_{xx}} - rx\partial_x V + \frac{(r-b)^2 x^2 (\partial_x V)^2}{\sigma^2 x^2 \partial_{xx} V} \\ &= -rx\partial_x V + \frac{1}{2}\frac{(r-b)^2 x^2 (\partial_x V)^2}{\sigma^2 x^2 \partial_{xx}}.\end{aligned}$$

Due to the good choice of the cost functional we find for $\alpha \in (0,1)$ a solution satisfying the HJB equation and having the correct final value to be

$$V(t,x) = e^{\lambda(T-t)} x^\alpha \qquad \text{with} \qquad \lambda = r\alpha + \frac{(b-r)^2 \alpha}{2\sigma^2 (1-\alpha)}.$$

This yields the optimal feedback function

$$u^*(x,t) = \frac{b-r}{\sigma^2(1-\alpha)}.$$

Hence, if $u^* \in [0,1]$ is valid, we have found the optimal strategy just to have a constant fraction of the wealth invested in both assets. Some special choices of the parameters make the optimal choice clearer: for $b \downarrow r$ we will not invest in the risky asset because it does not offer a higher average yield, for $\sigma \to \infty$ the same phenomenon occurs due to the concavity of the utility function penalizing relative losses higher than gains, for $\sigma \to 0$ or $\alpha \to 1$ we do not run into high risk when investing in the stock and thus will do so (even with borrowing for $u^* > 1$!),.

Bibliography

DURRETT, R. (1996): *Stochastic calculus. A practical introduction.* Probability and Stochastics Series. Boca Raton, FL: CRC Press.

FELLER, W. (1971): *An introduction to probability theory and its applications. Vol II. 2nd ed.* Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley and Sons.

KALLENBERG, O. (2002): *Foundations of modern probability. 2nd ed.* Probability and Its Applications. New York, Springer.

KARATZAS, I., AND S. E. SHREVE (1991): *Brownian motion and stochastic calculus. 2nd ed.* Springer, New York.

KLOEDEN, P. E., AND E. PLATEN (1992): *Numerical solution of stochastic differential equations.* Applications of Mathematics. 23. Berlin: Springer-Verlag.

LIPTSER, R. S., AND A. N. SHIRYAEV (2001): *Statistics of random processes. 1: General theory. 2nd ed.* Springer, Berlin.

ØKSENDAL, B. (1998): *Stochastic differential equations. An introduction with applications. 5th ed.* Universitext. Berlin: Springer.

REVUZ, D., AND M. YOR (1999): *Continuous martingales and Brownian motion. 3rd ed.* Springer, Berlin.

YONG, J., AND X. Y. ZHOU (1999): *Stochastic controls. Hamiltonian systems and HJB equations.* Applications of Mathematics. 43. New York, NY: Springer.