Lecture Notes on Stochastic Analysis

Taught summer term 2000
By Peter Mörters

Universität Kaiserslautern

Contents

Preface

The present volume represents my notes designed for the lecture "Stochastic Analysis" taught in the summer term 2000 at Universität Kaiserslautern. The course was the second part in a series of two courses on probability theory, my notes on the first part are available as Volume 6 in the same lecture note series. The series of courses will be rounded up with a course on stochastic differential equations in the next winter term.

The material covers an introduction to stochastic integration and selected applications. This selection concentrates on applications to the potential theory of Brownian motion, probabilistic representation of the solutions of partial differential equations and distributions of special functionals of Brownian motion. Although roughly half the students in the course are specialising in financial mathematics, I have decided against presenting applications to problems of financial mathematics in the course. My main reason for doing this was that there was a parallel course on the stochastics of financial markets by Prof. Korn, which was attended by most of my audience.

The principal sources of the material of the course are mentioned at the end of the script, all these sources contain suitable material for further studies. As usual there is a warning to all those who intend to use this script as a *replacement* for a lecture: the script does not include a vital ingredient, the pictures. Also I do not think that it can suitably convey my personal enthusiasm for the subject. The lecture was accompanied by a series of tutorials, which are not represented here, but were vital for the understanding of the material. My thanks go to Jochen Blath, who did an excellent job in the tutorials, and to my patient audience.

Peter Mörters,
Kaiserslautern, 3rd July 2000.

Chapter 1

Martingales and local martingales

1.1 Definition and examples of martingales

In the lecture on probability theory we have successfully developed a theory of martingales in *discrete time*. This theory has turned out to be as elegant as useful and our aim in this lecture is to derive a corresponding theory for the continuous time case. The notion of *stochastic integral* will be the core of this programme, but before we define stochastic integrals, the continuous time equivalent of the martingale transform, we will use this chapter to discuss examples of the use of martingales and an extension of the concept, the *local martingales*.

Definition: Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space and $\{\mathcal{F}(t) : t \geq 0\}$ a collection of sub-σ-fields with $\mathcal{F}(s) \subset \mathcal{F}(t)$ for all $s \leq t$, in other words a *filtration*. A (continuous time) process $\{X(t) : t \geq 0\}$ is *adapted to the filtration* $\{\mathcal{F}(t) : t \geq 0\}$ if every random variable $X(t)$ is $\mathcal{F}(t)$-measurable. An adapted process $\{X(t)\}$ is called an $\{\mathcal{F}(t)\}$*-martingale* if $\mathbf{E}|X(t)| < \infty$ for all $t \geq 0$ and

$$\mathbf{E}\Big\{X(t)\Big|\mathcal{F}(s)\Big\} = X(s) \text{ almost surely, for all } s \leq t.$$

If the filtration is clear from the context we will simply call $\{X(t) : t \geq 0\}$ a *martingale* without explicit reference to the filtration used. Most of the filtrations of interest have the property of being *right continuous*, which means that

$$\mathcal{F}(t) = \bigcap_{\varepsilon > 0} \mathcal{F}(t + \varepsilon).$$

Every process induces a natural filtration, where $\mathcal{F}(t)$ is generated by the collection of random variables $\{X(s) : s \leq t\}$. However, this filtration is not necessarily right continuous.

Theorem 1.1 *Let $\{B(t) : t \geq 0\}$ be standard Brownian motion defined by means of the projections on the Wiener space $(C[0,\infty), \mathcal{A}, \mathbb{P})$. Then let $\{\mathcal{F}^0(t)\}$ be the natural filtration and let*

$$\mathcal{F}(t) = \bigcap_{\varepsilon > 0} \mathcal{F}^0(t + \varepsilon).$$

Then $\{\mathcal{F}(t)\}$ is a right-continuous filtration and Brownian motion is a martingale with respect to this filtration.

Proof: The properties of $\{\mathcal{F}(t)\}$ are already known. We use the weak Markov property, to see

$$\mathbf{E}\Big\{X(t)\Big|\mathcal{F}(s)\Big\} = \mathbf{E}_{X(s)}\{B(t-s)\} = X(s) \text{ almost surely},$$

for all $s \le t$. ■

Remark: Brownian motion is also a martingale with respect to $\{\mathcal{F}^0(t)\}$. Recall that a comparison of these two filtrations at $t = 0$ was the key to Blumenthal's 01-law.

Recall that a vital notion in discrete time martingale theory was the notion of *uniform integrability*. A martingale $\{X(t) : t \ge 0\}$ is called *uniformly integrable* if, for every $\varepsilon > 0$, there is a $K > 0$ such that

$$\sup_{t \ge 0} \int_{\{|X(t)|>K\}} |X(t)| \, d\mathbf{P} < \varepsilon \, .$$

Brownian motion is *not* uniformly integrable, because for every $K > 0$,

$$\int_{\{|B(t)|>K\}} |B(t)| \, d\mathbf{P} = \frac{2}{\sqrt{2\pi t}} \int_K^\infty x e^{-x^2/2t} \, dx = \sqrt{\frac{2}{\pi}} \sqrt{t} e^{-K^2/2t} \to \infty \text{ as } t \to \infty.$$

Let us look at some more examples of martingales.

Theorem 1.2 *Let $\{B(t) : t \ge 0\}$ be standard Brownian motion. Then $\{B^2(t) - t : t \ge 0\}$ is a martingale with respect to $\{\mathcal{F}(t)\}$.*

Proof: Clearly, $\mathbf{E}|B^2(t) - t| \le 2t$ and we calculate

$$\begin{aligned}
\mathbf{E}\{B(t)^2|\mathcal{F}(s)\} &= \mathbf{E}\{B(s)^2 + 2B(s)(B(t) - B(s)) + (B(t) - B(s))^2|\mathcal{F}(s)\} \\
&= B(s)^2 + 2B(s)\mathbf{E}\{B(t) - B(s)|\mathcal{F}(s)\} + \mathbf{E}\{(B(t) - B(s))^2|\mathcal{F}(s)\} \\
&= B(s)^2 + (t - s).
\end{aligned}$$

This implies $\mathbf{E}\{B(t)^2 - t|\mathcal{F}(s)\} = B(s)^2 - s$ almost surely. ■

Theorem 1.3 *Let $\{B(t) : t \ge 0\}$ be standard Brownian motion. Then $\{B^3(t) - 3tB(t) : t \ge 0\}$ and $\{B^4(t) - 6tB(t)^2 + 3t^2 : t \ge 0\}$ are martingales with respect to $\{\mathcal{F}(t)\}$.*

This will be proved as an exercise. Of course, it is natural to ask what is special about the expressions appearing in the theorem and whether this series of examples can be extended to higher powers of Brownian motion. This will hopefully become clearer later in the lecture, if you want to explore it now you might want to look for the definition of *Hermite polynomials*.

If you think the last two examples were artificial, you should wait until the next section, when we use them to derive information about some nontrivial functionals of Brownian motion.

Theorem 1.4 *Suppose that $X(t) = \exp(\sigma B(t) + \mu t)$. The geometric Brownian motion $\{X(t) : t \ge 0\}$ is a martingale with respect to $\{\mathcal{F}(t) : t \ge 0\}$ if and only if the drift and volatility satisfy $\mu = -\sigma^2/2$.*

Proof: Note that

$$\begin{aligned}\mathbf{E}\{\exp(\sigma B(t))\} &= \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} e^{\sigma x}e^{-x^2/2t}\,dx\\ &= \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} e^{-(x/\sqrt{2t}-\sigma\sqrt{2t}/2)^2+\sigma^2 t/2}\,dx\\ &= e^{\sigma^2 t/2}\,.\end{aligned}$$

Using this we get

$$\begin{aligned}\mathbf{E}\{\exp(\sigma B(t))\,|\,\mathcal{F}(s)\} &= \exp(\sigma B(s))\mathbf{E}\{\exp(\sigma(B(t)-B(s))\,|\,\mathcal{F}(s)\}\\ &= \exp(\sigma B(s))\exp(\sigma^2(t-s)/2),\end{aligned}$$

using that the increment $B(t)-B(s)$ is independent of $\mathcal{F}(s)$ and has a normal distribution with mean 0 and variance $t-s$. ■

1.2 The optional stopping theorem

Some results from discrete time martingale theory can be extended by approximation to the continuous time setting. In this section we develop this approach systematically, prove the continuous version of the optional stopping theorem and give some nice applications.

A random variable $S:\Omega\to[0,\infty]$ is called a *stopping time* with respect to the filtration $\{\mathcal{F}(t)\}$ if, for all $t\ge 0$, $\{S<t\}\in\mathcal{F}(t)$. Define $\mathcal{F}(S)$ to be the σ-field of all $A\in\mathcal{A}$ such that

$$A\cap\{S\le t\}\in\mathcal{F}(t)\text{ for all }t\ge 0\,.$$

A good picture one can have in mind is that $\mathcal{F}(t)$ represents the information available at time t and S is the time when we buy some item. S is a stopping time if we base the decision whether to buy just on the knowledge available at the time. If the filtration is right continuous, then we can equivalently require $\{S\le t\}\in\mathcal{F}(t)$ in the definition of S.

In order to translate results from discrete time martingales to continuous time martingales, the following lemmas are useful.

Lemma 1.5 *If S is a stopping time and $S_n=([2^n S]+1)/2^n$, where $[\cdot]$ denotes the integer part, then S_n is a stopping time and $S_n\downarrow S$ almost surely.*

This was proved in the previous lecture.

Lemma 1.6 (Dominated convergence) *Suppose $X_n\to X$ almost surely and $|X_n|\le Z$ for all n and some Z with $\mathbf{E}|Z|<\infty$.*

(i) If $\mathcal{F}(n)$ is increasing and the union generates $\mathcal{F}$, then

$$\lim_{n\to\infty}\mathbf{E}\Big\{X_n\Big|\mathcal{F}(n)\Big\}=\mathbf{E}\Big\{X\Big|\mathcal{F}\Big\}\text{ almost surely.}$$

(ii) If $\mathcal{F}(n)$ is decreasing and the intersection is $\mathcal{F}$, then

$$\lim_{n\to\infty} \mathbf{E}\Big\{X_n \Big| \mathcal{F}(n)\Big\} = \mathbf{E}\Big\{X \Big| \mathcal{F}\Big\} \text{ almost surely.}$$

Proof: We prove the upward direction, the downward direction is an exercise. Let $W_N = \sup\{|X_n - X_m| : n, m \geq N\}$. $W_N \leq 2Z$ is integrable and $Y_n = \mathbf{E}\{W_N | \mathcal{F}(n)\}$ is a uniformly integrable martingale. Hence,

$$\limsup_{n\to\infty} \mathbf{E}\Big\{|X_n - X| \Big| \mathcal{F}(n)\Big\} \leq \limsup_{n\to\infty} \mathbf{E}\Big\{W_N \Big| \mathcal{F}(n)\Big\} = \mathbf{E}\Big\{W_N \,|\, \mathcal{F}\Big\},$$

by Lévy's Upward Theorem. Now

$$\begin{aligned} &\limsup_{n\to\infty} \Big|\mathbf{E}\{X_n|\mathcal{F}(n)\} - \mathbf{E}\{X|\mathcal{F}(n)\}\Big| \\ &\quad\leq \limsup_{n\to\infty} \mathbf{E}\Big\{|X_n - X| \Big| \mathcal{F}(n)\Big\} \\ &\quad\leq \mathbf{E}\Big\{W_N \,|\, \mathcal{F}\Big\}. \end{aligned}$$

The right hand side converges, as $N \to \infty$, to 0 by dominated convergence. ■

Suppose now that $\{X(n)\}$ is a discrete time martingale and $S \leq T$ are stopping times. Then $\{X(T \wedge n)\}$ is also a martingale and, *if* it is uniformly integrable, then by the martingale convergence theorem,

$$\lim_{n\to\infty} X(T \wedge n) = X(T) \text{ almost surely and in the } L^1\text{-sense.}$$

Note that L^1-convergence follows from uniform integrability. This already implies

$$X(n) = \mathbf{E}\{X(T)|\mathcal{F}(n)\} \text{ almost surely, for all } n,$$

because all uniformly integrable random variables come from accumulating data about the limit random variable. Informally, one could plug $S = n$ into this equation to get

$$\mathbf{E}\{X(T)|\mathcal{F}(S)\} = X(S) \text{ almost surely,} \tag{1.1}$$

but this is of course not a rigourous argument. To see that (1.1) is nevertheless true, we check the definition of conditional expectations. Clearly, $X(S)$ is $\mathcal{F}(S)$-measurable, and if $A \in \mathcal{F}(S)$,

$$\begin{aligned} \int_A X(S)\, d\mathbf{P} &= \sum_{n=0}^{\infty} \int_{A\cap\{S=n\}} X(n)\, d\mathbf{P} \\ &= \sum_{n=0}^{\infty} \int_{A\cap\{S=n\}} \mathbf{E}\{X(T)|\mathcal{F}(n)\}\, d\mathbf{P} \\ &= \sum_{n=0}^{\infty} \int_{A\cap\{S=n\}} X(T)\, d\mathbf{P} \\ &= \int_A X(T)\, d\mathbf{P}\,. \end{aligned}$$

In the penultimate step we have used that $A \cap \{S = n\} \in \mathcal{F}(n)$, by definition of $\mathcal{F}(S)$, and the definition of conditional expectation. This proves (1.1). This is a new variant of the optional stopping theorem, which we now generalise to the continuous time case.

Theorem 1.7 (Optional Stopping Theorem) *Suppose that $\{X(t) : t \geq 0\}$ is a continuous martingale and $S \leq T$ are stopping times such that $\{X(T \wedge t) : t \geq 0\}$ is uniformly integrable. Then*

$$\mathbf{E}\big\{X(T) \,|\, \mathcal{F}(S)\big\} = X(S),$$

and, in particular, $\{X(T \wedge t) : t \geq 0\}$ is a martingale with respect to $\{\mathcal{F}(T \wedge t) : t \geq 0\}$.

Remark: Note that T can be infinite with positive probability. Then $X(T)$ has to be interpreted as the limit of $X(T \wedge t)$ as $t \to \infty$, which exists by the martingale convergence theorem.

Proof: Define S_n corresponding to S as in the lemma. Then we can define a uniformly integrable discrete time martingale by putting

$$Y(m) = X(T \wedge m2^{-n}) \qquad , \qquad \mathcal{G}(m) = \mathcal{F}(m2^{-n})\,.$$

Now $\tilde{S} = 2^n S_n$ is an $\{\mathcal{G}(m)\}$-stopping time, because

$$\{\tilde{S} < m\} = \{S_n < m2^{-n}\} \in \mathcal{F}(m2^{-n}) = \mathcal{G}(m).$$

We can apply (1.1) to the stopping times $\tilde{S}$ and $\tilde{T} = \infty$. This implies

$$\mathbf{E}\{X(T)|\mathcal{F}(S_n)\} = \mathbf{E}\{Y(\tilde{T})|\mathcal{G}(\tilde{S})\} = Y(\tilde{S}) = X(S_n \wedge T)\,.$$

Now let $n \to \infty$ and apply continuity on the right and Lemma 1.6(ii) on the left hand side. The last remark follows by choosing $T \wedge s$ and $T \wedge t$, for $s < t$ as stopping times. ■

I now give a couple of my favourite applications of the optional stopping theorem and the fact that we know that the functionals of Brownian motion studied in the previous section are martingales.

Theorem 1.8 *Suppose $T = \inf\{t : B(t) \notin (-a, a)\}$ is the first exit time of standard Brownian motion from the interval $(-a, a)$. Then $\mathbf{E}\{T\} = a^2$ and $\mathrm{Var}\{T\} = 2a^4/3$.*

Proof: Applying the optional stopping theorem to the martingale $\{B^4(t) - 6tB(t)^2 + 3t^2 : t \geq 0\}$ and the stopping times $0 \leq T$, we get

$$\mathbf{E}\{B(T)^4 - 6TB(T)^2\} = -3\mathbf{E}\{T^2\}.$$

We know (and prove again below) that $\mathbf{E}\{T\} = a^2 < \infty$. Hence

$$a^4 - 6a^4 = -3\mathbf{E}\{T^2\}.$$

Solving this gives the result. ■

Theorem 1.9 *Suppose $T = \inf\{t : B(t) \notin (-a, a)\}$ is the first exit time of standard Brownian motion from the interval $(-a, a)$. Then, for all $\lambda \geq 0$,*

$$\mathbf{E}\Big\{\exp(-\lambda T)\Big\} = \frac{1}{\cosh(a\sqrt{2\lambda})}.$$

Proof: Applying the optional stopping theorem to the martingale $X(t) = \exp(\sigma B(t) - (\sigma^2/2)t)$ gives

$$1 = \mathbf{E}\Big\{ \exp\Big(\sigma B(T) - (\sigma^2/2)T\Big)\Big\}.$$

By symmetry, $B(T)$ and T are independent and $B(T)$ is uniformly distributed on $\{-a, a\}$. Hence

$$1 = \mathbf{E}\Big\{ \exp\Big(\sigma B(T) - (\sigma^2/2)T\Big)\Big\} = \frac{e^{\sigma a} + e^{-\sigma a}}{2}\,\mathbf{E}\Big\{ \exp\Big(-(\sigma^2/2)T\Big)\Big\}.$$

Now let $\sigma = \sqrt{2\lambda}$ and recall $\cosh(x) = (e^x + e^{-x})/2$. ■

The expression calculated in Theorem 1.9 is the *Laplace transform* of the nonnegative random variable T. The distribution of T is determined by means of the Laplace transform. Even if the Laplace transform cannot be inverted explicitly, one can derive important properties of the distribution from it. For example, we can recover the moments of T by differentiating the Laplace transform at 0 with respect to λ, e.g.

$$\mathbf{E}\{T\} = -\frac{\partial}{\partial\lambda}\mathbf{E}\Big\{ \exp(-\lambda T)\Big\}\Big|_{\lambda=0} = -\frac{\partial}{\partial\lambda}\frac{1}{\cosh(a\sqrt{2\lambda})}\Big|_{\lambda=0} = a^2.$$

Here is one more EXAMPLE. Let $b > 0$ and let $\{B(t) - bt : t \geq 0\}$ be a Brownian motion with downward drift and $a > 0$ a barrier. What is the probability that the motion hits the barrier? Let $T := \inf\{t > 0 : B(t) = a + bt\}$ be the (possibly infinite) waiting time until the drifted Brownian motion hits the barrier. We look at the martingale in Theorem 1.4 with $\sigma = b + \sqrt{b^2 + 2\lambda}$. Then

$$\begin{aligned}\sigma B(t\wedge T) - \frac{\sigma^2}{2}(t\wedge T) &\leq \sigma(a + b(t\wedge T)) - \frac{\sigma^2}{2}(t\wedge T)\\ &= (b + \sqrt{b^2+2\lambda})a - \lambda(t\wedge T),\end{aligned}$$

which is bounded from infinity. Hence the stopping theorem can be applied and gives that

$$\begin{aligned}1 &= \mathbf{E}\Big\{ \exp\Big(\sigma B(T\wedge t) - \sigma^2/2(T\wedge t)\Big)\Big\}\\ &\longrightarrow \mathbf{E}\Big\{ \exp\Big(\sigma B(T) - (\sigma^2/2)T\Big)1_{\{T<\infty\}}\Big\}\\ &= \mathbf{E}\{\exp(\sigma a - \lambda T)\Big),\end{aligned}$$

This implies

$$\mathbf{E}\Big\{ \exp(-\lambda T)\Big\} = \exp\Big(-a(b + \sqrt{b^2+2\lambda})\Big),$$

where the left hand side is interpreted as 0 if $T = \infty$. This is again one of those useful Laplace transforms. Just note in passing that the case $b = 0$ gives the Laplace transform of the first hitting time of level a for a standard Brownian motion, which we have studied in connection with the reflection principle (in this case the Laplace transform can be inverted). But we can also read off the solution to our problem from this expression. The probability that the motion hits the barrier is

$$\begin{aligned}\mathbf{P}\{T<\infty\} &= \lim_{\lambda\to 0}\mathbf{E}\Big\{ \exp(-\lambda T)\Big\}\\ &= \lim_{\lambda\to 0}\exp(-a(b+\sqrt{b^2+2\lambda}))\\ &= \exp(-2ab).\end{aligned}$$

This answers the original question. Another useful calculation gives the expected maximum of a Brownian motion with downward drift. We have

$$\begin{aligned}\mathbf{E}\Big\{\max_{t\geq 0} B(t) - bt\Big\} &= \int_0^\infty \mathbf{P}\{B(t) - bt \text{ hits } a\}\,da \\ &= \int_0^\infty e^{-2ab}\,da = \frac{1}{2b}\,.\end{aligned}$$

As a moral of this section we keep in mind that knowing that a process is a martingale can already lead to interesting and very concrete results. We now have a further motivation to study stochastic integrals. Our hope is, of course, that they are martingales and that we can exploit this fact in a similar manner as we have done for the simpler examples of this section.

1.3 Definition and examples of local martingales

In this section we define local martingales. This is a more general notion than the notion of martingale and helps us extending the class of integrands in the stochastic integral without having to invest much more work. The essential idea of this extension is that we require that certain properties of a process (like the martingale property) need to hold only locally.

Definition
Let $\{X(t) : t \geq 0\}$ be an adapted process and T a stopping time with respect to the filtration $\{\mathcal{F}(t) : t \geq 0\}$. Define the *stopped process* $\{X^T(t) : t \geq 0\}$ by $X^T(t) = X(T \wedge t)$. The process $\{X(t) : t \geq 0\}$ is called a *local martingale* with respect to $\{\mathcal{F}(t) : t \geq 0\}$ if there exists a sequence

$$0 = T_0 \leq T_1 \leq \ldots \leq T_n \uparrow \infty$$

of stopping times such that $\{X^{T_n}(t) : t \geq 0\}$ is a martingale. We say that $\{T_n\}$ is a *reducing* sequence of stopping times.

Remarks: Every martingale is a local martingale, because every sequence of stopping times, which increases to infinity is reducing. In the definition we can equivalently require that $\{X^{T_n}(t) : t \geq 0\}$ is a martingale with respect to $\{\mathcal{F}(t \wedge T_n) : t \geq 0\}$ instead of $\{\mathcal{F}(t) : t \geq 0\}$. Show this as an exercise.

Example: We construct a local martingale $\{X(t) : t \geq 0\}$, which is not a martingale. Although we will not prove all the details, I hope you get the flavour.

We let a particle do a symmetric random walk, but in continuous time and the waiting times between the jumps are random times. Fix a probability distribution on the integers by denoting the weight at the integer n by $p_n > 0$. We will assume that

$$\sum_{n=-\infty}^{\infty} n^2 p_n < \infty\,,$$

which means that the p_n are decreasing rather rapidly at both ends of the sequence. Now the process starts in 0 and stays there for an exponentially distributed time T_0 with expectation p_0, i.e. $X(t) = 0$ for all $t \in [0, T_0)$. Then we flip a coin and move to $Y_1 = \pm 1$ with equal probability, and wait again for an exponential time T_1, which is independent of T_0 and has expectation p_{Y_1}. In other words $X(t) = Y_1$ for $t \in [T_0, T_0 + T_1)$.

Suppose now we have just jumped to a level n and it was our kth jump. Then we stay there for an (independent) exponentially distributed time T_k with expectation p_n, before making the next jump of height ± 1, chosen independently and with equal probability. This means, heuristically, that we spend a long time when we jump to levels near 0 and only a short time at levels away from 0.

Because the symmetric random walk returns to 0 infinitely often and the times spend there are an independent, identically distributed sequence of times with positive expectation, the total time spent in 0 is infinite (by the strong law of large numbers) and hence we have defined the process on the whole time axis. The process is indeed a local martingale, to see this formally let

$$S_k = T_0 + \cdots + T_{k-1}$$

be the time of the kth jump. $\{S_k\}$ is a sequence of stopping times increasing to infinity. Moreover if $Y_1, Y_2, \dots$ is the i.i.d. sequence of jump heights, then

$$X(t \wedge S_k) = \sum_{i=1}^{k} Y_i \mathbf{1}_{\{S_i \le t\}} .$$

For $s < t$ and $F \in \mathcal{F}(s)$, for the natural filtration, the event $F \cap \{s < S_i \le t\}$ is independent of Y_i. Thus

$$\int_F X(t \wedge S_k) - X(s \wedge S_k)\, d\mathbf{P} = \sum_{i=1}^{k} \mathbf{E}\Big\{ Y_i \mathbf{1}_{\{s<S_i\le t\}\cap F} \Big\} = 0 ,$$

hence $\mathbf{E}\{X(t \wedge S_k)\,|\mathcal{F}(s)\} = X(s \wedge S_k)$, proving the martingale property of X^{S_k}. We give a heuristic argument why $\{X(t)\}$ is not a martingale: Let $s < t$ and suppose we are given the information that $X(s)$ has a very large value. If the process was a martingale, then this large value would be the expected value of $X(t)$. But it is not, because we know that the process spends only a small amount of time at the high values and, in fact, most of the time is spent near zero, so that the expected value for $X(t)$ given the unusual information about $X(s)$ is below the value of $X(s)$. A rigourous argument (using Markov chain theory) can be found in (4.2.6) in v. Weizsäcker/Winkler.

The examples of local martingales might not convince you that this class contains natural examples which fail to be martingales, but the next theorem shows a remarkable advantage of working with local martingales: If we look at *continuous* local martingales, we get uniform integrability for free. For example, it is worth considering Brownian motion as a local martingale and use a reducing sequence, such that the stopped Brownian motions are even uniformly integrable martingales.

Theorem 1.10 *Suppose $\{X(t) \,:\, t \ge 0\}$ is a continuous local martingale. Then the sequence*

$$T_n = \inf\{t \ge 0 \,:\, |X(t)| > n\}$$

is always reducing. In particular, we can find a sequence, which reduces $\{X(t)\}$ to a bounded (or uniformly integrable) martingale.

Proof: Suppose $0 < s < t$. If $\{S_n\}$ is a reducing sequence, then we apply the optional stopping theorem to $\{X^{S_n}(t)\}$ at times $s \wedge T_m$ and $t \wedge T_m$ and obtain, using the first remark of this section,

$$\mathbf{E}\{X(t \wedge T_m \wedge S_n) | \mathcal{F}(s \wedge T_m \wedge S_n)\} = X(s \wedge T_m \wedge S_n).$$

Multiplying by $1_{\{T_m > 0, S_n > 0\}} \in \mathcal{F}(s \wedge T_m \wedge S_n)$ we get

$$\mathbf{E}\{X(t \wedge T_m \wedge S_n) 1_{\{S_n > 0, T_m > 0\}} | \mathcal{F}(s \wedge T_m \wedge S_n)\} = X(s \wedge T_m \wedge S_n) 1_{\{S_n > 0, T_m > 0\}}.$$

As $n \to \infty$, $\mathcal{F}(s \wedge T_m \wedge S_n) \uparrow \mathcal{F}(s \wedge T_m)$ and

$$X(r \wedge T_m \wedge S_n) 1_{\{S_n > 0, T_m > 0\}} \to X(r \wedge T_m) 1_{\{T_m > 0\}},$$

for all $r > 0$. Because the sequence is dominated by m we get for the conditional expectations, using Lemma 1.6,

$$\mathbf{E}\Big\{X(t \wedge T_m) 1_{\{T_m > 0\}} \Big| \mathcal{F}(s \wedge T_m)\Big\} = X(s \wedge T_m) 1_{\{T_m > 0\}} \text{ almost surely,}$$

which proves the statement. ■

Remark: The proof also works for every sequence S_n of stopping times, which increases to infinity, but is smaller than T_n.

One more advantage of local martingales in comparison to martingales is that they can be defined easily on random time intervals $[0, \tau)$. Whereas the concept of martingale on $[0, \tau)$ is meaningless, because for large t the random variable $X(t)$ is not defined on the whole space Ω, the following definition of a local martingale on $[0, \tau)$ is very natural:

Definition
Suppose τ is a random time. The process $\{X(t) : t \in [0, \tau)\}$ is a *local martingale* if there exists a sequence $T_n \uparrow \tau$ of stopping times such that $\{X^{T_n}(t) : t \geq 0\}$ is a martingale.

Let us now explore the relationship between local martingales and martingales. We show that we can change the time of a local martingale so that we get a martingale. Recall our example, which is constructed by taking a martingale (symmetric random walk) and distorting the time scale, so that the martingale property is violated, but the local martingale property still holds.

Theorem 1.11 *Suppose $\{X(t) : t \in [0, \tau)\}$ is a continuous local martingale on an arbitrary, possibly random, time interval. Then there exists a time change $\gamma : [0, \infty) \to [0, \tau)$ such that each $\gamma(t)$ is a stopping time and the process $\{X(\gamma(t)) : t \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}(\gamma(t))\}$.*

Proof: If T_n are the reducing stopping times as in Theorem 1.10, then define $\gamma : [0, \infty) \to [0, \tau)$ by

$$\gamma(t) = \begin{cases} t - (k-1) & \text{if } T_{k-1} + (k-1) \leq t \leq T_k + (k-1) \\ T_k & \text{if } T_k + (k-1) \leq t \leq T_k + k. \end{cases}$$

Now optional stopping comes into play. Let $n = [t] + 1$. As $\gamma(t) \le T_n \wedge t$ we get $X(\gamma(t)) = \mathbf{E}\{X(T_n \wedge n)|\mathcal{F}(\gamma(t))\}$ and hence

$$\mathbf{E}\{X(\gamma(t))|\mathcal{F}(\gamma(s))\} = \mathbf{E}\Big\{X(T_n \wedge n)|\mathcal{F}(\gamma(s))\Big\} = X(\gamma(s)) \text{ almost surely,}$$

as desired. ■

One hopes that integrability conditions ensure that local martingales are martingales. The following is a positive result in this direction.

Theorem 1.12 *Suppose $\{X(t) : t \ge 0\}$ is a local martingale and, for every $t > 0$,*

$$\mathbb{E}\Big\{\sup_{0 \le s \le t} |X(s)|\Big\} < \infty\,,$$

then $\{X(t) : t \ge 0\}$ is a martingale.

Proof: Clearly, $\mathbf{E}|X(t)| < \infty$. Now, if $\{T_n\}$ is a reducing sequence,

$$\mathbb{E}\Big\{X^{T_n}(t)\,\Big|\mathcal{F}(s \wedge T_n)\Big\} = X^{T_n}(s) \text{ almost surely,}$$

by our assumption we can let $n \to \infty$ and use Lemma 1.6(i), observing that our condition makes sure that $X^{T_n}(t)$ is dominated by an integrable function. The limiting equation is the martingale property of $\{X(t)\}$. ■

From this we easily get the following important corollary.

Corollary 1.13 *A bounded local martingale is a martingale.*

Chapter 2

The stochastic integral

2.1 Predictable processes

Recall the discrete time situation and suppose that $\{X_n : n \geq 0\}$ is a martingale. Consider this as the price of an item of some good and let $\{C_n\}$ be the number of items you possess in the period $[n-1, n)$. As you can buy or sell items depending on the current price, C_n can change in time, but your decision to buy or sell at time $n-1$ can make use only of the price process up to time $n-1$, resp. the information available at time $n-1$. More formally, C_n must be $\mathcal{F}(n-1)$-measurable. Such a process is called *previsible* or sometimes *predictable.* The total profit you make by buying or selling according to the previsible process $\{C_n\}$ with a price process $\{X_n : n \geq 0\}$ is

$$(C \bullet X)_n = \sum_{k=1}^{n} C_k (X_k - X_{k-1}),$$

which is called *the martingale transform* of $\{X_n\}$ by $\{C_n\}$. Can you make a profit (on average) by choosing a clever strategy? The answer is no, as the following theorem says.

Theorem 2.1 (You can't beat the system) *If $\{X_n\}$ is a martingale and $\{C_n\}$ a previsible process, such that each C_n is bounded, then $\{(C \bullet X)_n\}$ is a martingale, and its expectation is* 0.

Proof: It is clear that $(C \bullet X)_n$ is integrable (as C_n is bounded) and by definition the process $\{(C \bullet X)_n\}$ is adapted to the filtration $\{\mathcal{F}(n)\}$ and starts in 0. Then, almost surely,

$$\mathbf{E}\{(C \bullet X)_n | \mathcal{F}(n-1)\} = \mathbf{E}\Big\{ \sum_{k=1}^{n} C_k (X_k - X_{k-1}) \Big| \mathcal{F}(n-1) \Big\}$$

$$= \sum_{k=1}^{n-1} C_k (X_k - X_{k-1}) + C_n \mathbf{E}\{X_n - X_{n-1} | \mathcal{F}(n-1)\} = (C \bullet X)_{n-1}.$$

This proves it all. ■

What is almost trivial in the discrete situation turns out to be rather deep (and accordingly also even more useful) in the continuous time setting. Roughly the first half of this course will be devoted to finding a continuous time analogue of this theorem. In the second half we will exploit the discoveries we have made in this search. In order to get the analogue, we have to

- find a suitable replacement for the previsible processes $\{C_n\}$. This will be the predictable processes defined in this section.
- find a suitable replacement for the martingale transform. This is the stochastic (Itô-) integral, we define in the latter parts of this chapter. The predictable processes will be the integrands of this integral.
- study the properties of the integral compared to the properties of classical integrals (e.g. Riemann integrals).

In this section we define an appropriate class of *functions that depend only on the past*, which we want to use as integrand for the stochastic integrals. As past and future are not as easily separated in continuous time as in discrete time, this is more subtle. The following example shows that it does not suffice to require $\{C(t)\}$ to be adapted, as one might perhaps suspect.

EXAMPLE: Define a process $\{X(t) : t \geq 0\}$ as follows. Let T be uniformly distributed on $[0,1]$ and independently $X = \pm 1$ with equal probability. Define

$$X(t) = 1_{\{T \leq t\}} X \, .$$

Let $\mathcal{F}(t)$ be the σ-field generated by $1_{\{T \leq t\}} X$. Then $\{X(t) : t \geq 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}(t)\}$. Our total profit if $\{X(t)\}$ is the price process and, at the same time, our buying strategy is $X(T)X = X^2 = 1$ and this is an almost surely positive profit.

The example just means that in order to get a reasonable theory we have to impose stronger constraints on the possible buying strategies, or equivalently integrands in the stochastic integral. A problem with the process in that the *right-continuity* of $X(t)$ allowed us to know at the time T already what the outcome of X is. If we exchange the $\leq$ for an $<$, in other words if we make the process *left continuous* we avoid this problem. This implies, for example, that we cannot react instantaneously to take advantage of a jump in the process we are betting on.

The simplest left-continuous integrand we can imagine is found by picking $a < b$ and $C \in \mathcal{F}(a)$ and setting

$$H(s, \omega) = C(\omega) 1_{(a,b]}(s). \tag{2.1}$$

In words, we buy $C(\omega)$ shares of stock at time a based on our knowledge at that time and sell them all at time b. Such integrands must be allowed in every reasonable theory and one should also be able to add finitely many of them and take limits.

Definition
On $[0, \infty) \times \Omega$ we define Π to be the σ-field generated by sets of the form $(a, b] \times A$, for $A \in \mathcal{F}(a)$. Π is called the *predictable σ-field*. A process $C : [0, \infty) \times \Omega \to \mathbf{R}$ (or $\mathbf{R}^d$) is called *predictable* if it is measurable with respect to the predictable σ-field.

Theorem 2.2 *The smallest σ-field on $[0, \infty) \times \Omega$, which makes all left-continuous adapted processes measurable is the predictable σ-field. In particular, all left-continuous, adapted processes —amongst them Brownian motion— are predictable.*

Proof: Denote the σ-field on $[0,\infty)\times\Omega$ generated by the left-continuous adapted processes by Π'. Trivially, $\Pi\subset\Pi'$, because the indicators of the events that generate Π are left-continuous, adapted processes. Conversely, let $H:[0,\infty)\times\Omega\to\mathbf{R}$ be left-continuous and adapted. Define $H^n(s,\omega)=H(m2^{-n},\omega)$ for $m2^{-n}<s\le(m+1)2^{-n}$. Then H^n is Π-measurable, because it is the sum of functions as in (2.1), and H^n converges to H, as H is left-continuous. ■

2.2 The variance process of a local martingale

Again we motivate the coming things by their discrete time analogue. Suppose $\{X_n\}$ is a martingale starting at $X_0=0$. Then, clearly, applying a nonlinear function f to the process leads to a process $\{f(X_n)\}$, which usually fails to be a martingale. If we find a process Y_n with some nice properties, such that $f(X_n)-Y_n$ is a martingale, we might still be able to make use of martingale theory in the analysis of the process $\{f(X_n)\}$. One can think of $\{Y_n\}$ as a *compensation*, which we can subtract from $\{f(X_n)\}$ to turn it into a martingale.

We now look at this problem in the case $f(x)=x^2$. Here the process we use as a compensation has the nice properties of being increasing and previsible. Note first (as an exercise) that, if $\{X_n\}$ is a martingale, then $\{X_n^2\}$ is still a submartingale, but not necessarily a martingale.

Theorem 2.3 (Doob decomposition) *Suppose $\{X_n\}$ is a martingale starting at $X_0=0$ with $\mathbf{E}\{X_n^2\}<\infty$ for all n. Then there is a unique previsible process $\{A_n\}$ such that $\{X_n^2-A_n\}$ is again a martingale starting in 0. Moreover, the process $\{A_n\}$ is increasing (in the weaker sense of being nondecreasing).*

Proof: Let $A_0=0$ and define, for $n\ge1$,

$$A_n=A_{n-1}+\mathbf{E}\{X_n^2|\mathcal{F}(n-1)\}-X_{n-1}^2.$$

From the definition one can see that A_n is $\{\mathcal{F}(n-1)\}$-measurable and, as $\{X_n^2\}$ is a submartingale, that $\{A_n\}$ is increasing. Also

$$\mathbf{E}\{X_n^2-A_n|\mathcal{F}(n-1)\}=\mathbf{E}\{X_n^2|\mathcal{F}(n-1)\}-A_n=X_{n-1}^2-A_{n-1},$$

hence the compensation is a martingale. To prove uniqueness assume that $\{A_n\}$ and $\{B_n\}$ both fulfil the requirements. Then $\{A_n-B_n\}$ is a previsible martingale starting at 0 and we infer that

$$A_n-B_n=\mathbf{E}\{A_n-B_n|\mathcal{F}(n-1)\}=A_{n-1}-B_{n-1}.$$

By induction we must have $A_n=B_n$. ■

The last argument in the proof indeed shows the following.

Corollary 2.4 *Every previsible martingale $\{X_n\}$ starting at $X_0=0$ is constant equal to 0.*

In the remainder of this section we will find a continuous analogue of these results. Later, we will also find nice compensations for all other smooth functions f in the continuous time setting. The answer will be Itô's formula.

Theorem 2.5 (Variance process) *If $\{X(t) : t \geq 0\}$ is a continuous local martingale with $X(0) = 0$, then there exists one and only one process $\langle X \rangle = \{\langle X \rangle_t : t \geq 0\}$, which is continuous, predictable and increasing, such that*

$$\{X(t)^2 - \langle X \rangle_t : t \geq 0\}$$

is a local martingale starting in 0. If X is a bounded martingale, then so is $X^2 - \langle X \rangle$.

Remarks:

a) Comparing with the discrete case observe that the concept of local martingales allowed us to get rid of the integrability condition.

b) The name of the process can be justified as follows. Suppose X is bounded, then

$$\mathrm{Var}\{X(t)\} = \mathbf{E}\{X(t)^2\} = \mathbf{E}\{\langle X \rangle_t\},$$

the variance of $X(t)$ is the expectation of $\langle X \rangle_t$. Sometimes, the variance process is also called the *increasing process of* $\{X(t)\}$, also for obvious reasons.

c) In the important special case of a Brownian motion the process $\{\langle B \rangle_t\}$ was already found in Theorem 1.2. Recall that $\{B(t)^2 - t\}$ is a martingale, so by the uniqueness part of the previous theorem we have

$$\langle B \rangle_t = t\,.$$

It is quite remarkable that in the case of Brownian motion this process is so simple, and in particular deterministic.

We shall now prove the **uniqueness part** of the theorem. We start the proof with the equivalent of Corollary 2.4, which will help us prove the uniqueness part. The straightforward translation of the corollary fails here, because there exist predictable, continuous martingales starting at 0, which are not 0. Think of Brownian motion.

Theorem 2.6 *Every continuous local martingale starting at 0, which is predictable and has bounded variation on every compact interval is constant equal to 0.*

Proof: Denote by

$$V(t) = \sup\Big\{\sum_{k=1}^{n} |X(t_k) - X(t_{k-1})| \,:\, 0 = t_0 < t_1 < \ldots < t_n = t\,,\ n \geq 1\Big\}$$

the *variation of X on [0,t].* If X is continuous and V locally bounded, then V is continuous. Define stopping times

$$S = S(k) = \inf\Big\{s \,:\, V(s) \geq k\Big\}.$$

If $t \leq S$, then $|X(t)| \leq k$. By Theorem 1.10 and the following remark, the sequence $\{S(k)\}$ is reducing and hence $M(t) = X(t \wedge S)$ is a bounded martingale with respect to $\{\mathcal{F}(t)\}$. A typical calculation shows

$$\begin{aligned}
\mathbf{E}\{(M(t) - M(s))^2 | \mathcal{F}(s)\} &= \mathbf{E}\{M(t)^2 | \mathcal{F}(s)\} - 2M(s)\mathbf{E}\{M(t)|\mathcal{F}(s)\} + M(s)^2 \\
&= \mathbf{E}\{M(t)^2 | \mathcal{F}(s)\} - M(s)^2 \\
&= \mathbf{E}\{M(t)^2 - M(s)^2 | \mathcal{F}(s)\}. \qquad (2.2)
\end{aligned}$$

Now let $0 = t_0 < t_1 < \ldots < t_n = t$ be a subdivision of $[0, t]$. Then

$$\begin{aligned}
\mathbf{E}\{M(t)^2\} &= \mathbf{E}\Big\{\sum_{m=1}^{n} M(t_m)^2 - M(t_{m-1})^2\Big\} \\
&= \mathbf{E}\Big\{\sum_{m=1}^{n} (M(t_m) - M(t_{m-1}))^2\Big\} \\
&\leq \mathbf{E}\Big\{V(t \wedge S) \sup_m |M(t_m) - M(t_{m-1})|\Big\} \\
&\leq k\mathbf{E}\Big\{\sup_m |M(t_m) - M(t_{m-1})|\Big\}.
\end{aligned}$$

If we now apply this to a sequence

$$\Delta_n = \{0 = t_0^n < t_1^n < \ldots < t_n^n = t\}$$

of partitions with mesh

$$\delta_n = \max_{i=1}^{n}(t_i^n - t_{i-1}^n)$$

going to 0, then continuity of M implies that

$$\lim_{n\to\infty} \sup_m |M(t_m^n) - M(t_{m-1}^n)| = 0.$$

The supremum is also bounded by $2k$, so bounded convergence yields that

$$\lim_{n\to\infty} \mathbf{E}\Big\{\sup_m |M(t_m^n) - M(t_{m-1}^n)|\Big\} = 0.$$

This proves that $\mathbf{E}\{M(t)^2\} = 0$, hence almost surely, $M(t) = 0$ for all rational $t \geq 0$, and by continuity this extends to all $t \geq 0$. Hence $\{X(t)\}$ is almost surely constant equal to 0. ■

To finish the proof of the uniqueness, suppose A and B are two processes with the desired properties. A and B are locally of bounded variation, as they are increasing. Now the process

$$\{A(t) - B(t) \,:\, t \geq 0\}$$

is still locally of bounded variation, continuous and a local martingale starting at 0, so it must be zero. This proves uniqueness.

Now we prove the **existence part**, which is considerably harder. We first show that there are continuous, predictable processes $\{Q(t)\}$, which make $\{X(t)^2 - Q(t)\}$ a martingale and are "almost increasing".

Lemma 2.7 *Let $\{X(t) \,:\, t \geq 0\}$ be a bounded, continuous martingale and $\{t_n\}$ a finite or infinite, strictly increasing sequence with $t_0 = 0$, which in the infinite case tends to ∞. Let $k(t) = \sup\{k \,:\, t_k < t\}$ and define*

$$Q(t) = \sum_{k=1}^{k(t)} (X(t_k) - X(t_{k-1}))^2 + (X(t) - X(t_{k(t)}))^2.$$

Then $\{X(t)^2 - Q(t)\}$ is a martingale starting in 0.

We call the process Q defined in the lemma the *Q-process* associated with X and the sequence $\{t_n\}$. The idea of our proof already becomes clearer from this lemma. Without the last term in the sum the process Q would be increasing. The effect of the last term gets smaller, if the points of $\{t_n\}$ are close together, so one can hope to get the desired object by taking a limit. Let us first prove the lemma.

Proof: Our standard calculation shows that if $r \le s \le t$, then

$$\begin{aligned}
&\mathbf{E}\{(X(t)-X(r))^2|\mathcal{F}(s)\} \\
&= \mathbf{E}\{(X(t)-X(s))^2|\mathcal{F}(s)\} + 2(X(s)-X(r))\mathbf{E}\{X(t)-X(s)|\mathcal{F}(s)\} + (X(s)-X(r))^2 \\
&= \mathbf{E}\{(X(t)-X(s))^2|\mathcal{F}(s)\} + (X(s)-X(r))^2.
\end{aligned}$$

Now, if $s < t$,

$$Q(t)-Q(s) = \Big(X(t)-X(t_{k(t)})\Big)^2 - \Big(X(s)-X(t_{k(s)})\Big)^2 + \sum_{k=k(s)+1}^{k(t)} (X(t_k)-X(t_{k-1}))^2.$$

Now we take conditional expectations and apply (2.2) as well as the calculation above with $r = t_{k(s)}$ and $t = t_{k(s)+1}$. We collect the terms, abbreviating $u(k(t)+1) = t$ and $u(i) = t_i$ for all $k(s) \le i \le k(t)$. Then

$$\begin{aligned}
&\mathbb{E}\{X(t)^2 - Q(t)\,|\,\mathcal{F}(s)\} \\
&= \mathbb{E}\{X(t)^2|\mathcal{F}(s)\} - Q(s) - \mathbb{E}\Big\{\sum_{i=k(s)+1}^{k(t)+1} \Big(X(u(i)) - X(u(i-1))\Big)^2 \Big|\mathcal{F}(s)\Big\} \\
&\quad + \mathbf{E}\{(X(s)-X(t_{k(s)}))^2|\mathcal{F}(s)\} \\
&= \mathbb{E}\{X(t)^2|\mathcal{F}(s)\} - Q(s) - \mathbb{E}\Big\{\sum_{i=k(s)+2}^{k(t)+1} X(u(i))^2 - X(u(i-1))^2 \Big|\mathcal{F}(s)\Big\} \\
&\quad - \mathbf{E}\{(X(t_{k(s)+1}) - X(s))^2\,|\,\mathcal{F}(s)\} - (X(s)-X(t_{k(s)}))^2 + (X(s)-X(t_{k(s)}))^2 \\
&= X(s)^2 - Q(s).
\end{aligned}$$

This finishes the proof of the lemma. ■

The key part of the proof is the following lemma, which was proved for the special case of Brownian motion in the last chapter of the previous lecture. If you look at the proof there, you can see that it used special properties of the normal distribution, so that it is no surprise that the general statement is harder to prove.

Lemma 2.8 *Let $\{X(t) : t \ge 0\}$ be a bounded, continuous martingale and*

$$\Delta_n = \{0 = t_0^n < t_1^n < \ldots < t_{N(n)}^n = t\}$$

a sequence of partitions of $[0,t]$ with mesh

$$\delta_n = \max_{i=1}^{N(n)} (t_i^n - t_{i-1}^n) \to 0.$$

Then there is a random variable A such that

$$\lim_{n\to\infty}\sum_{i=1}^{N(n)}(X(t_i^n)-X(t_{i-1}^n))^2 = A \text{ in the } L^2\text{-sense.}$$

The limit A is the *quadratic variation* of X on the interval $[0,t]$. The idea described after the formulation of the previous lemma indicates that this limit might be the right choice for $\langle X\rangle$. Once we have proved this lemma, we are almost done with the proof of the theorem. Let us begin with the proof.

Observation: $\mathbf{E}\{Q(t)-Q(s)|\mathcal{F}(s)\} = \mathbf{E}\{(X(t)-X(s))^2|\mathcal{F}(s)\}$.

Proof of the observation: Use Lemma 2.7, take conditional expectation and apply the argument of (2.2). This gives

$$\begin{aligned}\mathbf{E}\{Q(t)-Q(s)|\mathcal{F}(s)\} &= \mathbf{E}\{X(t)^2-X(s)^2|\mathcal{F}(s)\}\\ &= \mathbf{E}\{(X(t)-X(s))^2|\mathcal{F}(s)\}.\end{aligned}$$

The major technical step in the proof of the lemma is the following estimate.

Estimate: If $|X(t)|\le M$ for all t, then

$$\mathbf{E}\Big\{\Big(\sum_{k=1}^{N(n)}(X(t_i^n)-X(t_{i-1}^n))^2\Big)^2\Big\}\le 12M^4.$$

Proof of the estimate: For $0\le r\le t$ recall that $k(r)=\sup\{k: t_k^n<r\}$ and

$$Q(r)=\sum_{i=1}^{k(r)}(X(t_i^n)-X(t_{i-1}^n))^2+(X(r)-X(t_{k(r)}))^2.$$

Then, omitting the reference to n,

$$\begin{aligned}Q(t)^2 &= \Big(\sum_{m=1}^{N}\Big(X(t_m)-X(t_{m-1})\Big)^2\Big)^2\\ &= \sum_{m=1}^{N}\Big(X(t_m)-X(t_{m-1})\Big)^4+2\sum_{m=1}^{N-1}\Big(X(t_m)-X(t_{m-1})\Big)^2(Q(t)-Q(t_m)).\end{aligned}$$

Now we bound the expectation of the first term, using (2.2),

$$\begin{aligned}\mathbf{E}\Big\{\sum_{m=1}^{N}\Big(X(t_m)-X(t_{m-1})\Big)^4\Big\} &\le (2M)^2\mathbf{E}\Big\{\sum_{m=1}^{N}\Big(X(t_m)-X(t_{m-1})\Big)^2\Big\}\\ &= 4M^2\mathbf{E}\Big\{\sum_{m=1}^{N}X(t_m)^2-X(t_{m-1})^2\Big\}\\ &\le 4M^2\mathbf{E}\{X(t)^2\}\le 4M^4.\end{aligned}$$

To bound the second term we use the observation, and obtain

$$\begin{aligned}
&\mathbb{E}\Big\{\Big(X(t_m)-X(t_{m-1})\Big)^2(Q(t)-Q(t_m))\,\Big|\mathcal{F}(t_m)\Big\}\\
\le\ & \Big(X(t_m)-X(t_{m-1})\Big)^2\mathbf{E}\Big\{(Q(t)-Q(t_m))\,\Big|\mathcal{F}(t_m)\Big\}\\
=\ & \Big(X(t_m)-X(t_{m-1})\Big)^2\mathbf{E}\Big\{(X(t)-X(t_m))^2|\mathcal{F}(t_m)\Big\}\\
\le\ & (2M)^2(X(t_m)-X(t_{m-1}))^2.
\end{aligned}$$

Now we take expected value, sum over m, and recall (2.2), which gives

$$\begin{aligned}
&\mathbf{E}\Big\{\sum_{m=1}^{N-1}\Big(X(t_m)-X(t_{m-1})\Big)^2(Q(t)-Q(t_m))\Big\}\\
\le\ & 4M^2\mathbf{E}\Big\{\sum_{m=1}^{N-1}\Big(X(t_m)-X(t_{m-1})\Big)^2\Big\}\\
=\ & 4M^2\mathbf{E}\Big\{\sum_{m=1}^{N-1}X(t_m)^2-X(t_{m-1})^2\Big\}\\
\le\ & 4M^2\mathbf{E}\{X(t)^2\}\le 4M^4.
\end{aligned}$$

Putting all this together gives the estimate.

Proof of Lemma 2.8: Let P and Q be the Q-processes associated with $\{X(t)\}$ and two partitions of $[0,t]$. We have to show that $\mathbf{E}\{(P(t)-Q(t))^2\}$ converges to 0 as the mesh of these partitions goes to 0. Then our sequence is a Cauchy sequence in L^2 and, by completeness, we can infer the existence of a limit in the L^2-sense.

Applying the previous lemma twice and taking the difference shows that $Y(s)=P(s)-Q(s)$ defines a martingale. Let R be the Q-process associated with the martingale Y and the joint refinement of the two partitions. Hence, by the previous lemma, $\{Y^2(s)-R(s)\}$ is a martingale started in 0 and we infer

$$\mathbf{E}\{(P(t)-Q(t))^2\}=\mathbf{E}\{Y(t)^2\}=\mathbf{E}\{R(t)\}.$$

As $(a-b)^2\le 2a^2+2b^2$, for all real a,b, we have

$$R(t)\le 2P'(t)+2Q'(t)\,,$$

where now P' and Q' are the Q-processes associated with the process P and Q and the joint refinement of the partitions. It thus suffices (by symmetry) to show that $\mathbf{E}\{P'(t)\}$ converges to zero, when the mesh of the two partitions goes to zero.

To prove this let s_k be a partition point in the joint refinement and find a partition point t_j in the first partition such that

$$t_j\le s_k<s_{k+1}\le t_{j+1}.$$

Recalling the definition of the Q-process we get

$$\begin{aligned}
P(s_{k+1})-P(s_k) &= (X(s_{k+1})-X(t_j))^2-(X(s_k)-X(t_j))^2\\
&= (X(s_{k+1})-X(s_k))^2+2(X(s_{k+1})-X(s_k))(X(s_k)-X(t_j))\\
&= (X(s_{k+1})-X(s_k))\Big(X(s_{k+1})+X(s_k)-2X(t_j)\Big).
\end{aligned}$$

Now let S be the Q-process associated with $\{X(t)\}$ and the joint refinement. Summing squares of the previous expression yields

$$P'(t) \le S(t) \sup_k \Big\{X(s_{k+1}) + X(s_k) - 2X(t_{j(k)})\Big\}^2,$$

where $j(k) = \sup\{j : t_j \le s_k\}$. Applying the expected value and the Cauchy-Schwarz inequality gives

$$\mathbf{E}\{P'(t)\} \le \Big(\mathbf{E}\{S^2(t)\}\Big)^{1/2} \Big(\mathbf{E}\Big\{\sup_k\{X(s_{k+1}) + X(s_k) - 2X(t_{j(k)})\}^4\Big\}\Big)^{1/2}.$$

Here the first factor is bounded by $\sqrt{12}M^2$ by the estimate and the second factor goes to 0 whenever the mesh of both partitions goes to 0, by continuity of $\{X(t)\}$. This proves Lemma 2.8.

To prove the existence of the variance process we need some more tools. Both of them are of independent interest and the first one will be proved as an exercise.

Lemma 2.9 (Doob's L^2-maximal inequality) *If $\{X(t) : t \ge 0\}$ is a martingale, then*

$$\mathbf{E}\Big\{\sup_{s\le t} X(s)^2\Big\} \le 4\mathbf{E}\{X(t)^2\}.$$

Lemma 2.10 *Suppose for each n is $\{Z_n(t) : t \ge 0\}$ a martingale with respect to the filtration $\{\mathcal{F}(t)\}$ and, for each t, $Z_n(t) \to Z(t)$ in the L^p-sense, for some $p \ge 1$. Then $\{Z(t) : t \ge 0\}$ is a martingale.*

Proof: As convergence in the L^1-sense is the weakest, we may assume $p = 1$. We obtain

$$\begin{aligned}
\mathbf{E}\Big|\mathbf{E}\{Z_n(t)|\mathcal{F}(s)\} - \mathbf{E}\{Z(t)|\mathcal{F}(s)\}\Big| &= \mathbf{E}\Big|\mathbf{E}\{Z_n(t) - Z(t)|\mathcal{F}(s)\}\Big| \\
&\le \mathbf{E}\Big\{\mathbf{E}\Big\{|Z_n(t) - Z(t)|\,\Big|\,\mathcal{F}(s)\Big\}\Big\} \\
&= \mathbf{E}|Z_n(t) - Z(t)| \longrightarrow 0.
\end{aligned}$$

Now we can let $n \to \infty$ in $\mathbf{E}\{Z_n(t)|\mathcal{F}(s)\} = Z_n(s)$ and obtain by L^1-convergence the martingale property of $\{Z(t)\}$. ∎

Proof of the existence when $\{X(t)\}$ is a bounded martingale: Now let

$$\Delta_n = \{k2^{-n}t : 0 \le k \le 2^n\}.$$

Write P_n for the Q-process of the bounded martingale $\{X(t)\}$ associated with Δ_n. By our first lemma $\{P_n(t) - P_m(t)\}$ is a martingale. Using the L^2-maximal inequality we get

$$\mathbf{E}\Big\{\sup_{r\le t} |P_n(r) - P_m(r)|^2\Big\} \le 4\mathbf{E}\{|P_n(t) - P_m(t)|^2\}.$$

Because $\{P_m(t)\}$ is an L^2-Cauchy sequence we can pick an increasing sequence $n(k)$ such that for all $m \ge n(k)$,

$$\mathbf{E}\{|P_m(t) - P_{n(k)}(t)|^2\} \le 2^{-k}.$$

It follows from the last two observations and Markov's inequality that

$$\mathsf{P}\Big\{\sup_{r\le t}|P_{n(k)}(r)-P_m(r)|^2>1/k^2\Big\}\le k^2\mathsf{E}\Big\{\sup_{r\le t}|P_{n(k)}(r)-P_m(r)|^2\Big\}\le 4k^2 2^{-k}.$$

The right hand side is summable in k and hence, by Borel Cantelli, the event on the left hand side fails for at most finitely many k. Hence, there is a process $\{A(s)\}$ such that, almost surely,

$$\lim_{k\to\infty}\sup_{s\le t}|P_{n(k)}(s)-A(s)|=0\,.$$

Doing this for all intervals $[0,N]$ we can find a diagonal subsequence $n(k)$ such that, for all N,

$$\lim_{k\to\infty}\sup_{t\le N}|P_{n(k)}(t)-A(t)|=0\,.$$

Because each $\{P_n(t)\}$ is continuous, so is $\{A(t)\}$. The process is also predictable, because it is continuous and adapted. Furthermore, $\{A(t)\}$ is also increasing. Indeed, $P_m(t)$ is increasing when restricted to the points of Δ_m and hence $\{A(t)\}$ is increasing on a dense set and, by continuity, it is everywhere increasing. Finally, $\{X(t)^2-A(t)\}$ is a martingale by the last lemma.

To extend this result to local martingales we only need the following easy observation: The variance process of a stopped martingale does not increase after the stopping time. Formally this is:

Lemma 2.11 *If $\{X(t)\,:\,t\ge 0\}$ is a martingale and T a stopping time, then $\langle X^T\rangle=\langle X\rangle^T$.*

Proof: By the optional stopping theorem $(X^2)^T-\langle X\rangle^T$ is a local martingale and, clearly, $(X^2)^T=(X^T)^2$, so the statement follows from uniqueness. ■

Proof of the existence when $\{X(t)\}$ is a local martingale: Let T_n be a sequence reducing $\{X(t)\}$ to a bounded martingale $\{X^{T_n}(t)\}$. By the previous step there is a unique continuous, predictable, increasing process $\{A^n(t)\}$, which makes $\{(X^{T_n}(t))^2-A^n(t)\}$ a martingale. By the last lemma

$$A^n(t)=A^{n+1}(t)\text{ for all }t\le T_n.$$

Hence one can define

$$\langle X\rangle_t=A^n(t)\text{ for all }t\le T_n.$$

Then all the stated properties of $\{\langle X\rangle_t\}$ can be directly inferred from $\{A^n(t)\}$. This completes the proof of the existence of the variance process.

Note that we not only proved abstract existence of the variance process but also the following concrete limit representation.

Theorem 2.12 *Suppose that $\{X(t)\,:\,t\ge 0\}$ is a continuous local martingale. For every sequence $\{\Delta_n\}$ of subdivisions of $[0,t]$ with mesh going to 0, the Q-process converges uniformly in probability to the variance process, which means, for every $\varepsilon>0$,*

$$\lim_{n\to\infty}\mathsf{P}\Big\{\sup_{s\le t}|Q(s)-\langle X\rangle_s|>\varepsilon\Big\}=0.$$

Proof: Let $\delta, \varepsilon > 0$. Find a stopping time S such that $\{X^S(t)\}$ is a bounded martingale and $\mathbf{P}\{S \leq t\} \leq \delta$. For every subdivision the Q-processes Q and R of $\{X(t)\}$ and $\{X^S(t)\}$ agree on $[0, S]$ and so do $\{\langle X \rangle_t\}$ and $\{\langle X^S \rangle_t\}$. Now

$$\mathbf{P}\Big\{\sup_{s \leq t} |Q(s) - \langle X \rangle_s| > \varepsilon\Big\} \leq \delta + \mathbf{P}\Big\{\sup_{s \leq t} |R(s) - \langle X^S \rangle_s| > \varepsilon\Big\},$$

and by Lemma 2.8 the last term goes to 0 as the mesh goes to 0. ■

Suppose $\{X(t) : t \geq 0\}$ and $\{Y(t) : t \geq 0\}$ are continuous, local martingales. Then, in general, their product $XY = \{X(t)Y(t) : t \geq 0\}$ fails to be a local martingale. We now show that we can subtract a nice process and make it a local martingale again.

Theorem 2.13 *Suppose $X = \{X(t) : t \geq 0\}$ and $Y = \{Y(t) : t \geq 0\}$ are continuous, local martingales starting in 0. Then there is a unique continuous predictable process $\langle X, Y \rangle = \{\langle X, Y \rangle_t : t \geq 0\}$ that has bounded variation on every compact interval and starts in 0, such that $XY - \langle X, Y \rangle$ is a local martingale.*

The process $\langle X, Y \rangle$ is called the *covariance process* of X and Y. The **proof** of the theorem is easy. For **uniqueness** observe that, if A and B are two processes satisfying the requirements, then

$$A(t) - B(t) = (X(t)Y(t) - B(t)) - (X(t)Y(t) - A(t))$$

is a continuous predictable local martingale, which has bounded variation on every compact interval. By Theorem 2.6 we have $A - B = 0$.

For **existence** we define

$$\langle X, Y \rangle_t = \frac{1}{4}\Big(\langle X + Y \rangle_t - \langle X - Y \rangle_t\Big).$$

Then $\langle X, Y \rangle$ is a continuous predictable process and — as the difference of two increasing processes — it is of bounded variation on every compact interval. Finally,

$$X(t)Y(t) - \langle X, Y \rangle_t = \frac{1}{4}\Big((X(t) + Y(t))^2 - \langle X + Y \rangle_t - ((X(t) - Y(t))^2 - \langle X - Y \rangle_t)\Big)$$

is a local martingale. ■

2.3 Integration with respect to bounded martingales

We now assume that $\{X(t) : t \geq 0\}$ is a bounded continuous martingale and $\{H(t) : t \geq 0\}$ a predictable process and define the stochastic integral

$$\int_0^t H(s)\, dX(s) \text{ or } (H \cdot X)_t\,.$$

Similar as in the Lebesgue theory, we start with the definition for simple integrands and generalise step by step.

Step 1: Simple integrands.

We say that a stochastic process $\{H(t,\omega) : t \geq 0\}$ is a *simple predictable process* if, for some $t_0 < \cdots < t_m$ and $\mathcal{F}(t_{i-1})$-measurable random variables C_i,

$$H(t,\omega) = \sum_{i=1}^{m} 1_{(t_{i-1},t_i]}\, C_i \,.$$

The set of all simple predictable processes is denoted Π_1. If X is a continuous martingale and $t_0 \leq t \leq t_m$, we let

$$\int_0^t H(s)\, dX(s) := (H \cdot X)_t := \sum_{i=1}^{j(t)} C_i(X(t_i) - X(t_{i-1})) + C_{j(t)+1}(X(t) - X(t_{j(t)}))\,,$$

where

$$j(t) = \max\{k \in \{0, \ldots, m-1\} : t_k < t\}\,.$$

Also define

$$\int_0^t H(s)\, dX(s) := (H \cdot X)_t := \begin{cases} (H \cdot X)_{t_m} & \text{if } t \geq t_m \\ 0 & \text{if } t \leq t_0 \end{cases} .$$

This definition is very natural. Note that the choice of the $t_0 < \cdots < t_m$ is not unique for a given H, but it is easy to check that the integral is independent of this choice. We have the following crucial property.

Theorem 2.14 *If X is a continuous martingale and H a bounded, simple predictable process, then $\{(H \cdot X)_t : t \geq 0\}$ is a continuous martingale.*

Proof: Continuity is obvious, as is adaptedness and integrability. We check the martingale property. Let $s < t$. Then we can choose $t_0 < \cdots < t_m$ such that $s \in (t_{i-1}, t_i]$ and $t \in (t_{j-1}, t_j]$ for some $0 \leq i < j \leq m$. Then, using the tower property of conditional expectation,

$$\begin{aligned}
&\mathbf{E}\{(H \cdot X)_t \,|\, \mathcal{F}(s)\} - (H \cdot X)_s = \mathbf{E}\{(H \cdot X)_t - (H \cdot X)_s \,|\, \mathcal{F}(s)\} \\
&= \mathbf{E}\Big\{ \sum_{k=i+1}^{j-1} C_k(X(t_k) - X(t_{k-1})) + C_i(X(t_i) - X(s)) + C_j(X(t) - X(t_{j-1})) \Big| \mathcal{F}(s) \Big\} \\
&= \sum_{k=i+1}^{j-1} \mathbf{E}\Big\{ C_k \mathbf{E}\{X(t_k) - X(t_{k-1}) \,|\, \mathcal{F}(t_{k-1})\} \Big| \mathcal{F}(s) \Big\} + \mathbf{E}\Big\{ C_i \mathbf{E}\{X(t_i) - X(s) \,|\, \mathcal{F}(t_i)\} \Big| \mathcal{F}(s) \Big\} \\
&\quad + \mathbf{E}\Big\{ C_j \mathbf{E}\{X(t) - X(t_{j-1}) \,|\, \mathcal{F}(t_{j-1})\} \Big| \mathcal{F}(s)\} \\
&= 0\,.
\end{aligned}$$

This proves that the stochastic integral is a martingale. ■

Theorem 2.15 *Suppose X and Y are continuous martingales and $H, K \in \Pi_1$, then*

$$((H + K) \cdot X)_t = (H \cdot X)_t + (K \cdot X)_t\,,$$

$$(H \cdot (X + Y))_t = (H \cdot X)_t + (H \cdot Y)_t\,.$$

Proof: We can choose the $t_0 < \cdots < t_m$ such that $t \in (t_{k-1}, t_k]$ and

$$H = \sum_{i=1}^{m} 1_{(t_{i-1},t_i]} C_i^1 \text{ and } K = \sum_{i=1}^{m} 1_{(t_{i-1},t_i]} C_i^2 .$$

Then both sides of the first equality are equal to

$$\sum_{i=1}^{k-1} (X(t_i) - X(t_{i-1}))\,(C_i^1 + C_i^2) + (X(t) - X(t_{k-1}))\,(C_k^1 + C_k^2)$$

and both sides of the second equality are

$$\sum_{i=1}^{k-1} ((X(t_i) - X(t_{i-1}) + (Y(t_i) - Y(t_{i-1}))\,C_i^1 + ((X(t) - X(t_{k-1}) + (Y(t) - Y(t_{k-1}))\,C_k^1 .$$

This proves the statement. ■

Theorem 2.16 *If X and Y are bounded continuous martingales and H, K are bounded, simple predictable processes. Then,*

$$\langle H \cdot X, K \cdot Y \rangle_t = \int_0^t H(s)K(s)\, d\langle X, Y \rangle_s .$$

Consequently,

$$\mathbf{E}\{(H \cdot X)_t (K \cdot Y)_t\} = \mathbf{E}\{\int_0^t H(s)K(s)\, d\langle X, Y \rangle_s\}$$

and

$$\mathbf{E}\{(H \cdot X)_t^2\} = \mathbf{E}\{\int_0^t H(s)^2\, d\langle X \rangle_s\} .$$

Remark: Note that the integrands with respect to the functions of bounded variation $d\langle X, Y \rangle$ or $d\langle X \rangle_s$ are Riemann-Stieltjes integrals. In the present case the integrands are piecewise constant, so that the integrals are trivial.

This result will follow from the following lemma.

Lemma 2.17 *The process $Z = \{Z_t : t \geq 0\}$ defined by*

$$Z_t = (H \cdot X)_t (K \cdot Y)_t - \int_0^t H_s K_s\, d\langle X, Y \rangle_s$$

is a martingale.

Indeed, the main formula of the theorem says exactly this in terms of the definition of the covariance process, the second formula follows by taking expectations and the third by letting $H = K$, $X = Y$ and noting $\langle X, X \rangle = \langle X \rangle$. We now prove the lemma.

Proof: Both summands are additive in H resp. K (using the last theorem for the first summand) and hence it suffices to show the statement for the case $H = 1_{(a,b]} C$ and $K = 1_{(c,d]} D$, where the intervals $(a, b]$ and $(c, d]$ are either disjoint or identical.

In the *disjoint case* we have $H(s)K(s) = 0$ and hence we have to prove that $(H \cdot X)_t(K \cdot Y)_t$ is a martingale. To prove this let $J = C(X_b - X_a)D1_{(c,d]}$ and observe that

$$(J \cdot Y)_t = (H \cdot X)_t(K \cdot Y)_t$$

defines a martingale by Theorem 2.14.

In the *equality case* we have

$$Z_s = \begin{cases} 0 & \text{if } s \le a, \\ CD[(X(s) - X(a))(Y(s) - Y(a)) - (\langle X, Y\rangle_s - \langle X, Y\rangle_a)] & \text{if } a \le s \le b, \\ CD[(X(b) - X(a))(Y(b) - Y(a)) - (\langle X, Y\rangle_b - \langle X, Y\rangle_a)] & \text{if } s \ge b. \end{cases}$$

It suffices to check the martingale property for $a \le s < t \le b$. Note that $Z_t - Z_s$ equals

$$CD[X(t)Y(t) - X(s)Y(s) - X(a)(Y(t) - Y(s)) - Y(a)(X(t) - X(s)) - (\langle X, Y\rangle_t - \langle X, Y\rangle_s)].$$

Take conditional expectations and note that $X(a), Y(a)$ are $\mathcal{F}(s)$-measurable and $\mathbf{E}\{Y(t) - Y(s)|\mathcal{F}(s)\} = 0$. This leads to

$$\mathbf{E}\{Z_t - Z_s \,|\, \mathcal{F}(s)\} = CD\mathbf{E}\Big\{X(t)Y(t) - \langle X, Y\rangle_t - [X(s)Y(s) - \langle X, Y\rangle_s] \,\Big|\, \mathcal{F}(s)\Big\} = 0\,.$$

This finishes the proof of the lemma. ■

Step 2: Square integrable integrands.

We define $\Pi_2 = \Pi_2(X)$ to be the class of all predictable processes H, such that the norm $\|\cdot\|_X$ of H is finite. This norm is defined by

$$\|H\|_X := \Big(\mathbf{E}\int_0^\infty H_s^2\, d\langle X\rangle_s\Big)^{1/2}.$$

Let $\mathcal{M}^2$ be the set of all L^2-bounded martingales (with respect to the filtration $\{\mathcal{F}(t)\}$) with

$$\|X\|_2 := \Big(\sup_{t\ge 0} \mathbf{E}X_t^2\Big)^{1/2}.$$

These norms are introduced in such a way that the following programme works:

1) The stochastic integral (as defined so far) maps a subset of Π_2 isometrically into $\mathcal{M}^2$,

2) This subset, the bounded, simple predictable processes, is dense in Π_2,

3) $\mathcal{M}^2$ is complete, so that for every convergent sequence $H^n \to H$ in Π_2 the Cauchy sequence $(H^n \cdot X)$ in $\mathcal{M}^2$ has a limit in $\mathcal{M}^2$. This limit will be $(H \cdot X) \in \mathcal{M}^2$.

The first of our three steps is the following:

Theorem 2.18 *If $X \in \mathcal{M}^2$ is a continuous bounded martingale and $H \in \Pi_1$ is a bounded simple predictable process, then $\|H \cdot X\|_2 = \|H\|_X$.*

Proof: We have, by monotone convergence and Theorem 2.16,

$$\begin{aligned}\|H\|_X^2 &= \mathbf{E}\Big\{\int_0^\infty H_s^2\, d\langle X\rangle_s\Big\} = \sup_{t\geq 0}\mathbf{E}\Big\{\int_0^t H_s^2\, d\langle X\rangle_s\Big\} \\ &= \sup_{t\geq 0}\mathbf{E}\{(H\cdot X)_t^2\} = \|(H\cdot X)\|_2^2\,.\end{aligned}$$

■

Theorem 2.19 *If for $1\leq i\leq k$ we have $X^i\in\mathcal{M}^2$ and $H\in\Pi_2(X^i)$ then there is a sequence $\{H^n\}\subset\Pi_1$ of bounded simple predictable processes with*

$$\|H^n - H\|_{X^i}\longrightarrow 0 \textit{ for } 1\leq i\leq k.$$

Proof: If H is bounded, say by $M>0$, vanishes on (T,∞) and $X^i\in\mathcal{M}^2$, then we have

$$\|H\|_{X^i}^2 = \mathbf{E}\Big\{\int_0^\infty H_s^2\, d\langle X^i\rangle_s\Big\}\leq M^2\mathbf{E}\langle X^i\rangle_T\leq M^2\sup_{t\geq 0}\mathbf{E}\{(X_t^i)^2\}<\infty.$$

Let $\mathcal{H}_T$ be the collection of predictable H that vanish on (T,∞) for which the conclusion holds. We want to apply the Monotone Class Theorem (Lemma 4.6. in Probability Theory) to this class.

Recall that the predictable processes $H:[0,\infty)\times\Omega\to\mathbf{R}$ are the measurable functions with respect to the predictable σ-field, which is generated by the $\cap$-stable collection of events $(r,s]\times A$ with $A\in\mathcal{F}(a)$. Clearly, if $r<s\leq T$ and $A\in\mathcal{F}(r)$ then $H(t,\omega)=1_{(r,s]}(t)1_A(\omega)\in\mathcal{H}_T$.

Suppose now that $0\leq G^n\in\mathcal{H}_T$ and $G^n\uparrow G$ with G bounded. The dominated convergence theorem implies, as $n\to\infty$,

$$\|G-G^n\|_{X^i}^2 = \mathbf{E}\int (G_s - G_s^n)^2\, d\langle X^i\rangle_s\longrightarrow 0\,.$$

Hence we can pick n such that $\|G-G^n\|_{X^i}^2<\varepsilon^2$ for $1\leq i\leq k$. Since $G^n\in\mathcal{H}_T$ we can find a sequence $\{H^{n,m}\}$ of bounded simple predictable processes with

$$\|H^{n,m}-G^n\|_{X^i}\to 0 \text{ for } 1\leq i\leq k.$$

By the triangle inequality we can infer that $\|H^{n,m}-G\|_{X^i}<2\varepsilon$ for sufficiently large n,m, and hence $G\in\mathcal{H}_T$.

By the monotone class theorem $\mathcal{H}_T$ contains all bounded predictable processes that vanish on (T,∞). Now if $K\in\Pi_2(X^i)$ and we define $K^n = K1_{|K|\leq n}1_{[0,n]}$ then the dominated convergence theorem implies $\|K-K^n\|_{X^i}\to 0$. Since K^n is bounded and vanishes on (n,∞) we can see, by the triangle inequality, that K is a limit of bounded simple predictable processes. ■

Theorem 2.20 $\mathcal{M}^2$ *is complete.*

Proof: If $\mathcal{F}(\infty)$ is the σ-field generated by the union of all $\mathcal{F}(t)$ we construct a natural isometric isomorphism from $\mathcal{M}^2$ onto $L^2 = L^2(\Omega, \mathcal{F}(\infty), \mathbf{P})$, which we know is complete. The isometry is given by the martingale convergence theorem. Indeed, $\|X\|_2 < \infty$ implies that X is L^2-bounded and hence uniformly integrable. By martingale convergence theory there is a limit variable

$$X(\infty) = \lim_{s\to\infty} X(s) \text{ in } L^1.$$

We now prove that $X \mapsto X(\infty)$ is an isometric isomorphism from $\mathcal{M}^2$ to L^2.

Recall that —by uniform integrability— we have $X(t) = \mathbf{E}\{X(\infty) \,|\, \mathcal{F}(t)\}$. Suppose $Y \in L^2$ is given, then $Y(t) = \mathbf{E}\{Y|\mathcal{F}(t)\}$ defines a martingale with $Y(t) \to Y$ almost surely. By Jensen's inequality,

$$\mathbf{E}\{Y(t)^2\} = \mathbf{E}\{\mathbf{E}\{Y|\mathcal{F}(t)\}^2\} \le \mathbf{E}\{\mathbf{E}\{Y^2|\mathcal{F}(t)\}\} = \mathbf{E}\{Y^2\}. \tag{2.3}$$

We infer that $\{Y(t)\} \in \mathcal{M}^2$ and $Y \mapsto \{Y(t)\}$ is the inverse mapping to $X \mapsto X(\infty)$, which hence must be a bijection. To show that it is also an isometry we have to prove

$$\mathbf{E}\{X(\infty)^2\} = \sup_{t\ge 0} \mathbf{E}\{X(t)^2\}. \tag{2.4}$$

We have already seen in (2.3) that $\ge$ holds. To see the opposite direction we recall that L^2 boundedness of X implies that X^p is uniformly integrable for all $p < 2$. Hence, using martingale convergence and Jensen, for $p < 2$,

$$\mathbf{E}\{X(\infty)^p\} = \lim_{t\to\infty} \mathbf{E}\{X(t)^p\} \le \lim_{t\to\infty} \Big(\mathbf{E}\{X(t)^2\}\Big)^{p/2} \le \Big(\sup_{t\ge 0} \mathbf{E}\{X(t)^2\}\Big)^{p/2}.$$

Observe now that on the set $A = \{X(\infty) \le 1\}$ we have

$$\mathbf{E}\{1_A X(\infty)^2\} \le \mathbf{E}\{1_A X(\infty)^p\}$$

and on the set $B = \{X(\infty) \ge 1\}$ we have that $\{X(\infty)^p\}$ is increasing in p. Hence, when $p \uparrow 2$ by monotone convergence,

$$\limsup_{p\uparrow 2} \mathbf{E}\{X(\infty)^p\} \ge \mathbf{E}\{X(\infty)^2\}$$

and applying this limit to both sides in the previous inequality gives

$$\mathbf{E}\{X(\infty)^2\} \le \sup_{t\ge 0} \mathbf{E}\{X(t)^2\}.$$

This proves the isometry property (2.4). ■

We can now give the **definition of the stochastic integral** for square integrable integrands $H \in \Pi_2$ and bounded continuous martingales X as integrators. Let $\{H^n\}$ be a sequence of bounded simple predictable processes with

$$\|H^n - H\|_X \longrightarrow 0.$$

Since $\|H^n - H^m\|_X \to 0$ as $n, m \to \infty$ we get from the isometry property that

$$\|(H^n \cdot X) - (H^m \cdot X)\|_2 \to 0 \text{ as } n, m \to \infty.$$

In other words, $\{(H^n \cdot X)\}$ is a Cauchy sequence in $\mathcal{M}^2$ and thus —by completeness— must converge to a limit in $\mathcal{M}^2$. We define this limit to be the *stochastic integral*

$$(H \cdot X) =: \{(H \cdot X)_t \, : \, t \geq 0\} =: \Big\{ \int_0^t H(s) \, dX_s \, : \, t \geq 0 \Big\}.$$

The limit is automatically an L^2-bounded martingale, we just have to check it is well defined.

Lemma 2.21 *The definition of the stochastic integral $(H \cdot X)$ is independent of the choice of the approximating sequences $\{H^n\}$.*

Proof: Suppose $\{H^n\}$ and $\{\tilde{H}^n\}$ are two sequences of bounded simple predictable integrands with $\|H^n - H\|_X \to 0$ and $\|\tilde{H}^n - H\|_X \to 0$, let $H^n \cdot X \to Y$ and $\tilde{H}^n \cdot X \to \tilde{Y}$. Look at the sequence H_0^n with $H_0^{2n} = H^n$ and $H_0^{2n+1} = \tilde{H}^n$. Then $\|H_0^n - H\|_X \to 0$ and hence there is a $Z \in \mathcal{M}^2$ with $H_0^n \cdot X \to Z$. Now Z agrees with both Y and $\tilde{Y}$, hence they must coincide. ■

The major properties of the stochastic integral, in particular the fact that it is again a continuous martingale, is fortunately preserved by our approximation procedure.

Theorem 2.22 *If X is a bounded, continuous martingale and $H \in \Pi_2(X)$, then $H \cdot X \in \mathcal{M}^2$ is continuous.*

Proof: $H \cdot X \in \mathcal{M}^2$ is automatic from our definition. Choose an approximating sequence $\{H^n\}$ of bounded simple predictable processes and recall that $(H^n \cdot X)$ is a continuous martingale. Now we use Chebyshev's and Doob's L^2-maximal inequality to get

$$\begin{aligned} & \mathbb{P}\Big\{ \sup_{t \geq 0} |(H^n \cdot X)_t - (H \cdot X)_t| > \varepsilon \Big\} \\ \leq \;& \varepsilon^{-2} \mathbf{E}\Big\{ \sup_{t \geq 0} |(H^n \cdot X)_t - (H \cdot X)_t|^2 \Big\} \\ \leq \;& 4\varepsilon^{-2} \Big\| (H^n \cdot X) - (H \cdot X) \Big\|_2^2 \longrightarrow 0 \text{ as } n \to \infty \,. \end{aligned}$$

Now any sequence of integers going to ∞ has a subsequence $\{k_n\}$ along which almost surely,

$$\sup_{t \geq 0} |(H^{k_n} \cdot X)_t - (H \cdot X)_t| \to 0.$$

This implies that $(H \cdot X)$ is continuous. ■

The following isometry property follows easily from the definition of our integral.

Theorem 2.23 *If X is a bounded martingale and $H \in \Pi_2(X)$, then $\|H \cdot X\|_2 = \|H\|_X$.*

Proof: Choose an approximating sequence $\{H^n\}$ of bounded simple predictable processes and recall Lemma 2.18, to see

$$\|H \cdot X\|_2 = \lim_{n \to \infty} \|H^n \cdot X\|_2 = \lim_{n \to \infty} \|H^n\|_X = \|H\|_X \,.$$

■

We note that by approximation we easily see the linearity of the integral using Lemma 2.15.

Theorem 2.24 *If X is a bounded continuous martingale and $H, K \in \Pi_2(X)$, then $H + K \in \Pi_2(X)$ and we have*

$$\int_0^t (H(s) + K(s))\, dX_s = \int_0^t H(s)\, dX_s + \int_0^t K(s)\, dX_s\,.$$

Proof: Choose approximating sequences $\{H^n\}, \{K^n\}$ of bounded simple predictable processes and recall Lemma 2.15, to see

$$((H + K) \cdot X) = \lim_{n\to\infty}((H^n + K^n) \cdot X) = \lim_{n\to\infty}((H^n \cdot X) + (K^n \cdot X)) = (H \cdot X) + (K \cdot X).$$

■

2.4 The Kunita-Watanabe Inequality

In order to pass from bounded to local martingales as integrators in our stochastic integral and also to prove some key properties of the stochastic integrals we need a further tool, the Kunita-Watanabe Inequality. Note that none of the integrals involved in this formula is a stochastic integral.

Theorem 2.25 *Let X and Y be continuous local martingales and $H, K : \Omega \times [0, \infty) \to \mathbf{R}$ measurable processes. Let*

$$V(s) = \sup\Big\{ \sum_{k=1}^N \Big|\langle X, Y\rangle(t_k) - \langle X, Y\rangle(t_{k-1})\Big| \,:\, 0 = t_0 < \ldots < t_N = s\Big\}$$

be the total variation of the process $t \mapsto \langle X, Y\rangle_t$. Then, almost surely,

$$\int_0^t |H(s)K(s)|\, dV(s) \le \Big(\int_0^t H(s)^2\, d\langle X\rangle_s\Big)^{1/2} \Big(\int_0^t K(s)^2\, d\langle Y\rangle_s\Big)^{1/2}.$$

Remarks:

- The measurability of the processes refers to the product σ-field on $\Omega \times [0, \infty)$, which is much larger than the predictable σ-field.
- If X and Y are one and the same Brownian motion, then

  $$\langle X\rangle_s = \langle Y\rangle_s = \langle X, Y\rangle_s = V(s) = s\,.$$

  Then the Kunita-Watanabe Inequality is just the Cauchy-Schwarz Inequality.

Proof: Step 1: We use the abbreviation $\langle M, N\rangle_s^t = \langle M, N\rangle_t - \langle M, N\rangle_s$ and want to show that

$$(\langle X, Y\rangle_s^t)^2 \le \langle X, X\rangle_s^t \langle Y, Y\rangle_s^t\,.$$

First observe, for all $s \le t$ and λ,

$$\begin{aligned} 0 &\le \langle X+\lambda Y, X+\lambda Y\rangle_t - \langle X+\lambda Y, X+\lambda Y\rangle_s \\ &= \langle X,X\rangle_s^t + 2\lambda\langle X,Y\rangle_s^t + \lambda^2\langle Y,Y\rangle_s^t . \end{aligned}$$

This holds almost surely for all rational λ and all $s < t$. The nonnegative quadratic expression $a\lambda^2 + b\lambda + c$ has at most one real root, which means $b^2 - 4ac \le 0$, hence

$$(\langle X,Y\rangle_s^t)^2 \le \langle X,X\rangle_s^t \langle Y,Y\rangle_s^t .$$

Step 2: We prove Kunita-Watanabe for simple measurable integrands. For this purpose let $0 = t_0 < t_1 < \cdots < t_n = t$ be an increasing sequence of times, let h_i, k_i, $1 \le i \le n$ be bounded random variables, say $|h_i|, |k_i| \le M$ and define simple measurable processes

$$H(\omega, s) = \sum_{i=1}^n h_i(\omega) 1_{(t_{i-1}, t_i]}(s) , \quad K(\omega, s) = \sum_{i=1}^n k_i(\omega) 1_{(t_{i-1}, t_i]}(s) .$$

We define a refined partition

$$0 = s_0 < s_1 < \cdots < s_N = t$$

such that

$$V(t_i) - V(t_{i-1}) \le \sum_j |\langle X,Y\rangle_{s_j} - \langle X,Y\rangle_{s_{j-1}}| + \frac{\varepsilon}{nM^2} ,$$

where the sum extends over those j with $s_j \in (t_{i-1}, t_i]$. From the definition of the integral, the first step and Cauchy-Schwarz we infer

$$\begin{aligned} \int_0^t |H(s)K(s)|\, dV(s) &= \sum_{i=1}^n |h_i k_i| \,(V(t_i) - V(t_{i-1})) \\ &\le \sum_{i=1}^n |h_i k_i| \sum_j |\langle X,Y\rangle_{s_{j-1}}^{s_j}| + \varepsilon \\ &\le \sum_{i=1}^n \sum_j |h_i||k_i| \Big(\langle X,X\rangle_{s_{j-1}}^{s_j} \langle Y,Y\rangle_{s_{j-1}}^{s_j} \Big)^{1/2} + \varepsilon \\ &\le \Big(\sum_{i=1}^n h_i^2 \langle X,X\rangle_{t_{i-1}}^{t_i} \Big)^{1/2} \Big(\sum_{i=1}^n k_i^2 \langle Y,Y\rangle_{t_{i-1}}^{t_i} \Big)^{1/2} + \varepsilon . \end{aligned}$$

Noting that $\varepsilon > 0$ was arbitrary finishes the proof of the inequality for the case of simple measurable functions.

Step 3: Let $M > 0$ be large and introduce a stopping time

$$T = \inf \Big\{ t \,:\, \langle X\rangle_t \text{ or } \langle Y\rangle_t > M \Big\} .$$

By monotone convergence it suffices to prove the theorem for the case that $H = K = 0$ for $s \ge T$ and $|H_s|, |K_s| \le M$ for $s \le T$. Then on $[0, T]$ the increasing processes $\langle X\rangle$, $\langle Y\rangle$ induce finite measures of total mass $\le M$. Now for arbitrary measurable $H \ge 0$, bounded by M, there is a decreasing sequence $\{H^n\}$ of simple measurable processes with $H \le H^n \le M$, such that

$$\int_0^t (H_s^n)^2 \, d\langle X\rangle_s \longrightarrow \int_0^t (H_s)^2 \, d\langle X\rangle_s .$$

Together with the same result for K and Y we get the Kunita-Watanabe inequality for all bounded measurable processes. ■

We first apply the Kunita-Watanabe Inequality to generalise the statements of Theorems 2.15 and 2.16 from the case of simple integrands to integrands in $\Pi_2(X)$.

Theorem 2.26 *Suppose X and Y are bounded continuous martingales and $H \in \Pi_2(X) \cap \Pi_2(Y)$, then $H \in \Pi_2(X+Y)$ and*

$$(H \cdot (X+Y))_t = (H \cdot X)_t + (H \cdot Y)_t \,.$$

Proof: Note first that

$$\langle X+Y \rangle_t = \langle X \rangle_t + \langle Y \rangle_t + 2\langle X, Y \rangle_t$$

and the Kunita-Watanabe inequality and the inequality of the geometric and arithmetic mean imply that, for the variation V of $\langle X, Y \rangle$,

$$V(t) - V(s) \le (\langle X \rangle_s^t \langle Y \rangle_s^t)^{1/2} \le \frac{1}{2}(\langle X \rangle_s^t + \langle Y \rangle_s^t).$$

We infer that

$$\langle X+Y \rangle_s^t \le \langle X \rangle_s^t + \langle Y \rangle_s^t + 2(V(t) - V(s)) \le 2(\langle X \rangle_s^t + \langle Y \rangle_s^t).$$

This implies that $H \in \Pi_2(X+Y)$. To prove the formula note that by Theorem 2.19 we can find a sequence $\{H^n\}$ so that $\|H^n - H\|_Z \to 0$ for all $Z = X, Y, X+Y$. By Theorem 2.15 we have

$$(H^n \cdot (X+Y))_t = (H^n \cdot X)_t + (H^n \cdot Y)_t \,.$$

Now we can let $n \to \infty$ and use that $(H^n \cdot Z) \to (H \cdot Z)$ in $\mathcal{M}^2$ for all $Z = X, Y, X+Y$. ■

Theorem 2.27 *If X and Y are bounded continuous martingales and $H \in \Pi_2(X)$ and $K \in \Pi_2(Y)$, then*

$$\langle H \cdot X, K \cdot Y \rangle_t = \int_0^t H(s) K(s) \, d\langle X, Y \rangle_s \,.$$

Proof: As in Theorem 2.16 the statement means showing that the process $Z = \{Z_t : t \ge 0\}$ defined by

$$Z_t = (H \cdot X)_t (K \cdot Y)_t - \int_0^t H_s K_s \, d\langle X, Y \rangle_s$$

is a martingale. Let H^n and K^n be sequences of bounded simple predictable processes converging in $\Pi_2(X)$ resp. $\Pi_2(Y)$ to H and K, respectively. Let $Z^n = \{Z_t^n : t \ge 0\}$ the result when replacing H, K by H^n, K^n in the definition of Z. By Theorem 2.16 Z^n is a martingale and hence, by Lemma 2.10, Z is a martingale if Z_s^n converges in L^1 to Z_s. To show this, note that the triangle inequality and the linearity of the integral show that

$$\begin{aligned} &\mathbb{E}\Big\{ \sup_{t \ge 0} |(H^n \cdot X)_t (K^n \cdot Y)_t - (H \cdot X)_t (K \cdot Y)_t| \Big\} \\ &\le \mathbb{E}\Big\{ \sup_{t \ge 0} |((H^n - H) \cdot X)_t (K^n \cdot Y)_t| \Big\} + \mathbb{E}\Big\{ \sup_{t \ge 0} |(H \cdot X)_t ((K^n - K) \cdot Y)_t| \Big\}. \end{aligned}$$

To estimate the first term we use Cauchy-Schwarz, the L^2-maximal inequality and Theorem 2.23.

$$\begin{aligned}
&\mathbf{E}\Big\{\sup_{t\geq 0}|((H^n-H)\cdot X)_t(K^n\cdot Y)_t|\Big\} \\
&\leq \mathbf{E}\Big\{\sup_{t\geq 0}|((H^n-H)\cdot X)_t|\sup_{t\geq 0}|(K^n\cdot Y)_t|\Big\} \\
&\leq \Big(\mathbf{E}\Big\{\sup_{t\geq 0}|((H^n-H)\cdot X)_t|^2\Big\}\mathbf{E}\Big\{\sup_{t\geq 0}|(K^n\cdot Y)_t|^2\Big\}\Big)^{1/2} \\
&\leq 4\|((H^n-H)\cdot X)\|_2\|(K^n\cdot Y)\|_2 \\
&= 4\|H^n-H\|_X\|K^n\|_Y \longrightarrow 0.
\end{aligned}$$

The estimate of the other term is analogous. Now we have shown that $(H^n\cdot X)_t(K^n\cdot Y)_t$ converges to $(H\cdot X)_t(K\cdot Y)_t$ in L^1 and it remains to show convergence of the second term in the definition of Z^n resp. Z. To do this we have to recall the triangle inequality for Stieltjes integrals with respect to functions of bounded variation

$$\Big|\int_0^t f(s)dm(s)\Big| \leq \int_0^t |f(s)|dV(s)\,,$$

where $V(s)$ is the variation of the function m on the interval $[0,s]$. Using this and denoting V the variation of the function $\langle X,Y\rangle$, we get

$$\begin{aligned}
&\Big|\int_0^t H_s^nK_s^n\,d\langle X,Y\rangle_s-\int_0^t H_sK_s\,d\langle X,Y\rangle_s\Big| \\
&\leq \int_0^t |H_s^nK_s^n-H_sK_s|\,dV(s) \\
&\leq \int_0^t |H_s^n-H_s||K_s^n|\,dV(s)+\int_0^t |H_s||K_s^n-K_s|\,dV(s)\,.
\end{aligned}$$

Again it suffices to look at the first term. Using now Kunita-Watanabe, the first term is

$$\leq \Big(\int_0^t |H_s^n-H_s|^2\,d\langle X\rangle_s\Big)^{1/2}\Big(\int_0^t |K_s^n|^2\,d\langle Y\rangle_s\Big)^{1/2}.$$

Taking expected values and using the Cauchy-Schwarz inequality

$$\begin{aligned}
&\mathbf{E}\Big\{\int_0^t |H_s^n-H_s||K_s^n|\,dV(s)\Big\} \\
&\leq \Big(\mathbf{E}\int_0^t |H_s^n-H_s|^2\,d\langle X\rangle_s\Big)^{1/2}\Big(\mathbf{E}\int_0^t |K_s^n|^2\,d\langle Y\rangle_s\Big)^{1/2} \\
&\leq \|H^n-H\|_X\|K^n\|_Y\,,
\end{aligned}$$

which converges to 0. We thus have shown that $(H^n\cdot X)_t(K^n\cdot Y)_t$ converges to $(H\cdot X)_t(K\cdot Y)_t$ in L^1, which finishes the proof. ∎

2.5 Integration with respect to local martingales

We do the final step in the definition of the stochastic integral we generalise from bounded to local martingales as integrators and slightly enlarge the class of integrands to be

$$\Pi_3(X) := \Big\{ H \text{ predictable} : \int_0^t H_s^2 \, d\langle X\rangle_s < \infty \text{ almost surely for all } t \geq 0 \Big\} .$$

For the definition we use an almost obvious fact.

Lemma 2.28 *Suppose X, Y are bounded continuous martingales and $H, K \in \Pi_2(X)$.*

(1) *If for a stopping time T we have $H_s = K_s$ and $X_s = Y_s$ for all $s \leq T$. Then $(H \cdot X)_s = (K \cdot Y)_s$ for all $s \leq T$.*

(2) *If for two stopping times $S \leq T$ we have $H_s = K_s$ for all $S < s \leq T$ and $X_s - X_t = Y_s - Y_t$ for all $S \leq s, t \leq T$. Then, for all $S \leq s \leq T$,*

$$(H \cdot X)_s - (H \cdot X)_S = (K \cdot Y)_s - (K \cdot Y)_S .$$

Proof: Look at the first part and let $H_s^{(1)} = H_s 1_{\{s \leq T\}}$ and $K_s^{(1)} = K_s 1_{\{s \leq T\}}$ and write $H_s = H_s^{(1)} + H_s^{(2)}$ and $K_s = K_s^{(1)} + K_s^{(2)}$. Clearly, $(H^{(1)} \cdot X^T)_t = (K^{(1)} \cdot Y^T)_t$ for all $t \geq 0$. We also have

- $H_s^{(2)} = K_s^{(2)} = 0$ for all $s \leq T$,
- $X_s^T - X_s = Y_s^T - Y_s = 0$ for all $s \leq T$,

but it is not easy to prove without technical equipment that this implies $(H^{(2)} \cdot X)_t = (K^{(2)} \cdot Y)_t = 0$ and $(H^{(1)} \cdot (X - X^T))_t = (K^{(1)} \cdot (Y - Y^T))_t = 0$ for all $t \leq T$. To see this, we can use Theorem 2.27, which implies

$$0 \leq \langle H^{(2)} \cdot X\rangle_t = \int_0^t (H_s^{(2)})^2 \, d\langle X\rangle_s = 0 \text{ for all } t \leq T.$$

Now from Theorem 2.12 it is easy to infer that $(H^{(2)} \cdot X)_t = 0$ for $t \leq T$ almost surely. Essentially the same argument gives

$$0 \leq \langle H^{(1)} \cdot (X - X^T)\rangle_t = \int_0^t (H_s^{(1)})^2 \, d\langle X - X^T\rangle_s = 0 \text{ for all } t \leq T,$$

which implies $(H^{(1)} \cdot (X - X^T))_t = 0$ for $t \leq T$ almost surely. The three parts of the argument together with the linearity of the integral (Theorem 2.24 and 2.26) imply the first statement. For the second statement one can define $H_s^{(1)} = H_s 1_{\{S < s \leq T\}}$ and $K_s^{(1)} = K_s 1_{\{S < s \leq T\}}$. Then one observes that, trivially,

$$(H^{(1)} \cdot (X - X^S)^T)_t = (K^{(1)} \cdot (Y - Y^S)^T)_t$$

for all $t \geq 0$. The remaining terms

$$(H^{(2)} \cdot (X - X^S))_t \text{ and } (H^{(1)} \cdot (X^T - X))_t \text{ and } (H^{(1)} \cdot X^S)_t$$

can be estimated for $S \le t \le T$ with the same argument as in the first part. ■

Definition of the stochastic integral: Assume that X is a continuous local martingale and $H \in \Pi_3(X)$. Let $\{S_n\}$ be an increasing sequence of stopping times with

$$S_n \le T_n = \inf\Big\{t \,:\, |X_t| > n \text{ or } \int_0^t H_s^2 \, d\langle X\rangle_s > n\Big\}$$

and $S_n \uparrow \infty$. Define H^n by $H_s^n = H_s 1_{\{s\le S_n\}}$ and let X^n be X stopped in S_n. Observe that if $m < n$ Lemma 2.28 implies that $(H^m \cdot X^m)_s = (H^n \cdot X^n)_s$ for $s \le S_m$. Also $H^n \in \Pi_2(X^n)$. Hence we can define the stochastic integral $(H \cdot X)$ by

$$(H\cdot X)_t := \int_0^t H(s)\, dX_s := (H^n \cdot X^n)_t \text{ for } t \le S_n.$$

We have to check that this definition is independent of the choice of the stopping times $\{S_n\}$. Indeed, suppose $S_n, R_n \le T_n$ are sequences of stopping times going to infinity, let $H_s^n = H_s 1_{\{s\le S_n\}}$ and $\tilde H_s^n = H_s 1_{\{s\le R_n\}}$ and let X^n and $\tilde X^n$ be the martingales stopped accordingly. Then Lemma 2.28(i) implies that,

$$(\tilde H^n \cdot \tilde X^n)_s = (H^n \cdot X^n)_s \text{ for all } s \le S_n \wedge R_n,$$

i.e. the definitions coincide.

There are some immediate consequences of the definition, which we state now. A useful fact is that stopping the integral at a stopping time T, stopping the integrator at time T or setting the integrator $= 0$ after time T all lead to the same result.

Theorem 2.29 *Let S be a stopping time, X a continuous local martingale and $H \in \Pi_3(X)$. Define H^S by $H_s^S = H_s 1_{\{s\le S\}}$, then*

$$(H\cdot X)^S = (H^S \cdot X^S) = (H \cdot X^S) = (H^S \cdot X)\,.$$

Proof: For $s \le S$ all these integrals agree by the first part of Lemma 2.28. Further all these four processes stay constant after the stopping time S, by the second part of Lemma 2.28. ■

Theorem 2.30 *If X is a continuous local martingale and $H \in \Pi_3(X)$, then $(H \cdot X)$ is a continuous local martingale.*

Proof: By stopping at

$$T_n = \inf\{t\ge 0 \,:\, \int_0^t H_s^2 \, d\langle X\rangle_s > n \text{ or } |X_t| > n\}$$

we obtain, by Theorem 2.29 and Theorem 2.22, that

$$(H\cdot X)^{T_n} = (H^{T_n} \cdot X^{T_n})$$

is a continuous martingale. The result follows as $T_n \to \infty$ almost surely. ■

Theorem 2.31 *If X, Y are continuous local martingales and $H, K \in \Pi_3(X)$, then $H + K \in \Pi_3(X)$ and we have*

$$\int_0^t (H(s) + K(s))\, dX_s = \int_0^t H(s)\, dX_s + \int_0^t K(s)\, dX_s$$

and if also $H \in \Pi_3(Y)$ then $H \in \Pi_3(X+Y)$ and

$$\int_0^t H(s)\, d(X+Y)_s = \int_0^t H(s)\, dX_s + \int_0^t H(s)\, dY_s\,.$$

Proof: If $H, K \in \Pi_3(X)$, then use $(x+y)^2 \le 4x^2 + 4y^2$ to see that

$$\int_0^t (H_s + K_s)^2\, d\langle X\rangle_s \le 4\int_0^t H_s^2\, d\langle X\rangle_s + 4\int_0^t H_s^2\, d\langle X\rangle_s < \infty\,,$$

hence $H + K \in \Pi_3(X)$. Stopping at

$$S_n = \inf\{t \ge 0\, :\, \int_0^t H_s^2\, d\langle X\rangle_s > n \text{ or } \int_0^t K_s^2\, d\langle X\rangle_s > n \text{ or } |X_t| > n\}$$

makes sure that H^{S_n} and $K^{S_n} \in \Pi_2(X^{S_n})$ and X^n is a bounded martingale, so that the first part is reduced to Theorem 2.24. For the second result observe that $H \in \Pi_3(X+Y)$ follows from $\langle X+Y\rangle_s^t \le 2(\langle X\rangle_s^t + \langle Y\rangle_s^t$ as in Theorem 2.26. Stopping in

$$S_n = \inf\{t \ge 0\, :\, \int_0^t H_s^2\, d\langle X\rangle_s > n \text{ or } \int_0^t H_s^2\, d\langle Y\rangle_s > n \text{ or } |X_t| > n \text{ or } |Y_t| > n\}$$

reduces the result to Theorem 2.26. ∎

Theorem 2.32 *If X and Y are continuous local martingales and $H \in \Pi_3(X)$ and $K \in \Pi_3(Y)$, then*

$$\langle H\cdot X, K\cdot Y\rangle_t = \int_0^t H(s)K(s)\, d\langle X, Y\rangle_s\,.$$

Proof: Use the stopping times

$$S_n = \inf\{t \ge 0\, :\, \int_0^t H_s^2\, d\langle X\rangle_s > n \text{ or } \int_0^t K_s^2\, d\langle Y\rangle_s > n \text{ or } |X_t| > n \text{ or } |Y_t| > n\}$$

and Theorem 2.27 to infer that

$$\langle H^{S_n}\cdot X^{S_n}, K^{S_n}\cdot Y^{S_n}\rangle_t = \int_0^t H^{S_n}(s)K^{S_n}(s)\, d\langle X^{S_n}, Y^{S_n}\rangle_s\,.$$

Now from Theorem 2.12 we know that

$$\langle X, Y\rangle_s = \langle X^{S_n}, Y\rangle_s = \langle X^{S_n}, Y^{S_n}\rangle_s \text{ for } s \le S_n.$$

Using this and Theorem 2.29 we have

$$\langle H^{S_n}\cdot X^{S_n}, K^{S_n}\cdot Y^{S_n}\rangle_t = \langle (H\cdot X)^{S_n}, (K\cdot Y)^{S_n}\rangle_t = \langle H\cdot X, K\cdot Y\rangle_t \text{ for all } t \le S_n$$

and

$$\int_0^t H^{S_n}(s)K^{S_n}(s)\,d\langle X^{S_n},Y^{S_n}\rangle_s = \int_0^t H(s)K(s)\,d\langle X,Y\rangle_s \text{ for all } t \le S_n\,.$$

■

Finally, the following property, the *associative law* will be proved as an exercise. The basic technique is the same as in the previous cases, one has to proceed along the three steps of the construction of the stochastic integral.

Theorem 2.33 *If X is a continuous local martingale, $K \in \Pi_3(X)$ and $H \in \Pi_3(K \cdot X)$, then $HK \in \Pi_3(X)$ and*

$$H \cdot (K \cdot X) = ((HK) \cdot X) \text{ or equivalently}$$

$$\int_0^t H_s\,d(K \cdot X)_s = \int_0^t H_s K_s\,dX_s \text{ for all } t \ge 0\,.$$

We now give a representation of the integral which is reminiscent of the Riemann (or Stieltjes) integral. Note that this formula requires H to be continuous (this condition can be relaxed) and uses the *left endpoints* for the evaluation of the Riemann(-Stieltjes) sums, which is *essential.*

Theorem 2.34 *Let $\{\Delta_n\}$ be a sequence of partitions*

$$\Delta_n = \left\{0 = t_0^n < t_1^n < \cdots < t_{N(n)}^n = t\right\}$$

of $[0,t]$ with mesh $\delta_n = \sup|t_i^n - t_{i-1}^n| \to 0$. Suppose that X is a continuous local martingale and H a continuous predictable process. Then, in probability,

$$\lim_{n\to\infty} \sum_{i=1}^{N-1} H(t_i^n)(X(t_{i+1}^n) - X(t_i^n)) = \int_0^t H(s)\,dX_s\,.$$

For the proof we first need a lemma, which later be useful in another context.

Lemma 2.35 *Let X be a continuous local martingale with $\langle X\rangle_s \le M$ for all $s \ge 0$. If $\{H^n\}$ is a sequence of predictable processes and $|H_s^n| \le M$ for all s,ω, such that*

$$\sup_s |H_s^n - H_s| \longrightarrow 0 \text{ in probability,}$$

then $(H^n \cdot X) \to (H \cdot X) \in \mathcal{M}^2$.

Proof: Lemma 2.23 and the bounded convergence theorem imply

$$\|(H^n - H)\cdot X\|_2^2 = \|H^n - H\|_X^2 = \mathbf{E}\int_0^\infty (H^n - H)^2\,d\langle X\rangle_s \le M\mathbf{E}\Big\{\sup_s |H_s^n - H_s|^2\Big\} \longrightarrow 0\,,$$

as $n \to \infty$. ■

Proof of the Theorem: Observe that $H \in \Pi_3(X)$ so that the right hand side is well defined. We define a stopping time

$$T = \inf \Big\{ t \,:\, t, \langle X \rangle_t \text{ or } |H_s| > M \Big\}$$

and replace X by X^T and H by $H(\cdot \wedge T)$, which still is continuous. By Theorem 2.29 it suffices to prove the result for these stopped processes. We define approximations H^n of H by

$$H^n_s = H_{t^n_i} \text{ for } s \in (t^n_i, t^n_{i+1}]\,,$$

$H^n_s = H^n_t$ for $s > t$. Observe that, by the first step in the definition of stochastic integrals

$$\sum_{i=0}^{N(n)-1} H(t^n_i)(X(t^n_{i+1}) - X(t^n_i)) = (H^n \cdot X)_t\,.$$

Since $s \mapsto H_s$ is continuous we have for each ω that

$$\sup_{s \geq 0} |H^n_s - H_s| \longrightarrow 0\,.$$

Therefore our result follows directly from Lemma 2.5 upon observing that the $\mathcal{M}^2$-convergence of martingales obviously implies convergence in probability of the martingales at fixed times. ■

2.6 Semimartingales

For the formulation and proof of the next results it is useful to consider an approach that allows us to unify classical Stieltjes integration and stochastic integration.

Definition
A continuous process $X = \{X_t \,:\, t \geq 0\}$ is called a *semimartingale* if there is a continuous local martingale $M = \{M_t \,:\, t \geq 0\}$ and a continuous adapted process $A = \{A_t \,:\, t \geq 0\}$ started at 0, which is locally of bounded variation, such that $X_t = M_t + A_t$ for all $t \geq 0$.

Theorem 2.36 *If X is a continuous semimartingale, then the decomposition $X = M + A$ is unique.*

Proof: Suppose $X = M + A = M' + A'$. Then the process $A' - A = M - M'$ is a continuous local martingale, locally of bounded variation and starting in 0. Hence it is identically equal to 0, by Theorem 2.6. ■

Definition of integrals for semimartingales: Suppose $X = M + A$ is a continuous semimartingale. Define Π to be the class of *locally bounded predictable processes*, consisting of all those predictable processes H such that there exist an increasing sequence of stopping times $T_n \uparrow \infty$, such that$|H(\omega, s)| \leq n$ for all $s \leq T_n$. Note that $\Pi \subset \Pi_3(M)$ because

$$\int_0^{T_n} H_s^2 \, d\langle M \rangle_s \leq n^2 \langle M \rangle_{T_n} < \infty.$$

Hence we can define

$$(H \cdot X)_t := (H \cdot M)_t + \int_0^t H(\omega, s)\, dA_s\,,$$

where the first integral is a stochastic integral and the second is a Stieltjes integral. The following properties are all easy to obtain from our results for stochastic integrals and the analogous result for Stieltjes integrals.

- If $H \in \Pi$ and X is a continuous semimartingale, then the process $(H \cdot X)$ is also a continuous semimartingale.
- If $H, K \in \Pi$ and X, Y are continuous semimartingales, then the following rules hold
  - Distributive laws:
  $$((H + K) \cdot X) = (H \cdot X) + (K \cdot X)\,,$$
  $$(H \cdot (X + Y)) = (H \cdot X) + (H \cdot Y)\,.$$
  - Associative law:
  $$((HK) \cdot X) = (H \cdot (K \cdot X))\,.$$
- Define the covariance process $\langle X, Y\rangle$ of two semimartingales $X = M + A$ and $Y = N + B$ by
  $$\langle X, Y\rangle_t = \langle M, N\rangle_t.$$
  Then Theorem 2.32 can be written as $\langle H \cdot X, K \cdot Y\rangle = ((HK) \cdot \langle X, Y\rangle)$.

Chapter 3

Stochastic calculus and Itô's formula

In this chapter we derive the results which correspond to the fundamental theorem of calculus, which in the classical world relates differentiation and integration. Because we do not have the concept of "stochastic differentiation" we can only deal with those formulas, which do not explicitly mention differentiation.

3.1 The integration by parts formula

If $F, G : \mathsf{R} \to \mathsf{R}$ are differentiable functions with derivatives $f, g : \mathsf{R} \to \mathsf{R}$, then the integration by parts rule says that

$$\int_0^t f(s)G(s)\,ds = F(t)G(t) - F(0)G(0) - \int_0^t F(s)g(s)\,ds\,.$$

For Stieltjes integrals this formula can be represented without the use of derivatives, using that integration $f(s)\,ds = dF(s)$ and $g(s)\,ds = dG(s)$. Then we can write

$$\int_0^t G(s)\,dF(s) = F(t)G(t) - F(0)G(0) - \int_0^t F(s)\,dG(s)\,,$$

and the formula holds for all continuous F, G of bounded variation. We are looking for an analog of this formula for stochastic integration, but mind that the formula will have a different form now. We start with an easy generalisation of Theorem 2.32 for several semimartingales.

Lemma 3.1 *Let X^i and Y^j be continuous semimartingales and $H^i \in \Pi$, $K^j \in \Pi$. Define*

$$X = \sum_{i=1}^n (H^i \cdot X^i) \text{ and } Y = \sum_{j=1}^m (K^j \cdot Y^j).$$

Then

$$\langle X, Y\rangle_t = \sum_{i=1}^n \sum_{j=1}^m \int_0^t H_s^i K_s^j \, d\langle X^i, Y^j\rangle_s\,.$$

Proof: Recall that for $X = M + A$, $Y = N + B$ we have $\langle X, Y\rangle = \langle M, N\rangle$ by definition and that $\langle\,\cdot\,,\,\cdot\,\rangle$ is linear in both entries. The result follows from this together with Lemma 2.32. ■

We note an immediate consequence of Theorem 2.12 and the definition of the covariance process.

Lemma 3.2 *Suppose X and Y are continuous semimartingales and $\{\Delta_n\}$ a sequence of partitions of $[0,t]$ with mesh $\delta_n \to 0$. Then, in probability,*

$$\lim_{n\to\infty} \sum_{k=1}^{N} (X(t_k^n) - X(t_{k-1}^n))(Y(t_k^n) - Y(t_{k-1}^n)) = \langle X, Y\rangle_t\,.$$

Proof: For a given partition $0 = t_0 < \ldots < t_N = t$ and A continuous and of bounded variation on $[0,t]$ we have

$$\sum_{k=1}^{N} |(X(t_k^n) - X(t_{k-1}^n))(A(t_k^n) - A(t_{k-1}^n))| \le \sup_{|x-y|\le\delta_n} |X(x) - X(y)| V(t) \to 0 \text{ as } n \to 0\,.$$

Hence for all semimartingales $X = M + A$, $Y = N + B$, using Theorem 2.12,

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{k=1}^{N} (X(t_k^n) - X(t_{k-1}^n))(Y(t_k^n) - Y(t_{k-1}^n)) \\
&= \lim_{n\to\infty} \sum_{k=1}^{N} (M(t_k^n) - M(t_{k-1}^n))(N(t_k^n) - N(t_{k-1}^n)) \\
&= \lim_{n\to\infty} \frac{1}{4}\Big(\sum_{k=1}^{N} (M(t_k^n) + N(t_k^n) - M(t_{k-1}^n) - N(t_{k-1}^n))^2 \\
&\quad - \sum_{k=1}^{N} (M(t_k^n) - N(t_k^n) - M(t_{k-1}^n) + N(t_{k-1}^n))^2 \Big) \\
&= \frac{1}{4}(\langle M+N\rangle_t - \langle M-N\rangle_t) \\
&= \langle M, N\rangle_t = \langle X, Y\rangle_t.
\end{aligned}$$

■

The previous lemmas are the main ingredient in the proof of our integration by parts formula.

Theorem 3.3 (Integration by Parts) *If X and Y are continuous semimartingales, then*

$$\int_0^t Y(s)\,dX(s) = X(t)Y(t) - X(0)Y(0) - \int_0^t X(s)\,dY(s) - \langle X, Y\rangle_t\,.$$

Proof: Let $\{\Delta_n\}$ be a sequence of subdivisions of $[0,t]$ with mesh $\delta_n \to 0$. We can write

$$\begin{aligned}
X(t)Y(t) - X(0)Y(0) &= \sum_{i=0}^{N-1} (X(t_{i+1}^n) - X(t_i^n))(Y(t_{i+1}^n) - Y(t_i^n)) \\
&\quad + \sum_{i=0}^{N-1} Y(t_i^n)(X(t_{i+1}^n) - X(t_i^n)) + \sum_{i=0}^{N-1} X(t_i^n)(Y(t_{i+1}^n) - Y(t_i^n))\,.
\end{aligned}$$

By Lemma 3.2 we have

$$\sum_{i=0}^{N-1} (X(t_{i+1}^n) - X(t_i^n))(Y(t_{i+1}^n) - Y(t_i^n)) \longrightarrow \langle X, Y \rangle_t$$

and the last two terms converge by Theorem 2.34 and the definition of the Stieltjes integral to

$$\int_0^t Y_s \, dX_s + \int_0^t X_s \, dY_s \, .$$

This gives the desired result. ■

3.2 Itô's formula

Itô's formula now plays the role of the fundamental theorem of calculus, which you probably remember in the form

$$F(x(t)) - F(x(0)) = \int_0^t f(x(s))x'(s) \, ds \, .$$

Here $f = F'$ and x is differentiable. Writing this as a Stieltjes integral we get

$$F(x(t)) - F(x(0)) = \int_0^t f(x(s)) \, dx(s) \, ,$$

and this formula holds when x is continuous and of bounded variation. Itô's formula is an analog of this for the case that x is a local martingale. Then a third term enters, which makes the existence of a second derivative of F necessary.

Theorem 3.4 (Itô's formula) *Let $F : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable and X a continuous semimartingale. Then, almost surely, for all $t \geq 0$,*

$$F(X_t) - F(X_0) = \int_0^t F'(X_s) \, dX_s + \frac{1}{2} \int_0^t F''(X_s) \, d\langle X \rangle_s.$$

Remarks:

- Suppose that X is a local martingale starting in 0 and $F(0) = 0$. Then $F' \circ X \in \Pi_3(X)$ and we have represented the process $\{F(X_t) \, : \, t \geq 0\}$ by a local martingale plus a compensation, which is nice in the sense that it is of bounded variation. Recall that the case $F(x) = x^2$ was our motivation to introduce the variance process $\langle X \rangle$.
- If $F : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable and X a continuous semimartingale, then $\{F(X_t) \, : \, t \geq 0\}$ is also a continuous semimartingale.

Instead of proving the formula now, we prove directly a more powerful version, dealing with the case of several semimartingales — note that Theorem 3.4 is a special case of the following theorem. In the exercises an alternative (more transparent) proof will be discussed.

Theorem 3.5 (Itô's formula) *Let $f : \mathbf{R}^d \to \mathbf{R}$ be such that the partial derivatives*

$$\partial_j f = \frac{\partial f}{\partial x_j} \text{ and } \partial_{ij} f = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

exist for all $1 \le i,j \le d$ and are continuous. Let $X^1, \ldots, X^d$ be continuous semimartingales and write $X_t = (X_t^1, \ldots, X_t^d)$. Then

$$f(X_t) - f(X_0) = \sum_{i=1}^d \int_0^t \partial_i f(X_s)\, dX_s^i + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \int_0^t \partial_{ij} f(X_s)\, d\langle X^i, X^j\rangle_s.$$

Proof: By stopping, it suffices to prove the result when $|X_t^i|, \langle X^i\rangle_t \le M$ for all i, t. Any function f satisfying the hypotheses, can be approximated by polynomials g_n in such a way that $g_n, \partial_i g_n$ and $\partial_{ij} g_n$ converge uniformly on $[-M, M]^d$ to $f, \partial_i f, \partial_{ij} f$. If we now show the result for polynomials we can use Lemma 2.35 to pass from g_n to f.

It thus suffices to prove the result for polynomials —and by linearity even for monomials. So let

$$f(x) = x^{k_1} \cdots x^{k_n} \text{ for } k_1, \ldots, k_n \in \{1, \ldots, d\}.$$

Here we use superscripts, not powers, and allow $k_i = k_j$. If $n = 1$ Itô's formula reads

$$X_t^{k_1} - X_0^{k_1} = \int_0^t 1\, dX_s^{k_1},$$

which holds by definition. For larger n we use induction. Let

$$Y_t = \prod_{m=1}^n X_t^{k_m}$$

be a monomial for which the formula holds and let $Z_t = X_t^{k_{n+1}}$. Take a second to recall what the first and second derivatives of f look like. Our induction hypothesis hence is Itô's formula for $f(x) = x^{k_1} \cdots x^{k_n}$,

$$\begin{aligned} Y_t - Y_0 &= \sum_{i=1}^n \int_0^t \Big(\prod_{m=1, m\ne i}^n X_s^{k_m} \Big)\, dX_s^{k_i} \\ &\quad + \frac{1}{2} \sum_{i=1}^n \sum_{j=1, j\ne i}^n \int_0^t \Big(\prod_{m=1, m\ne i,j}^n X_s^{k_m} \Big)\, d\langle X^{k_i}, X^{k_j}\rangle_s. \end{aligned}$$

Note that the formula shows that Y is a semimartingale. By the integration by parts formula we have

$$Y_t Z_t - Y_0 Z_0 = \int_0^t Z_s\, dY_s + \int_0^t Y_s\, dZ_s + \langle Y, Z\rangle_t.$$

Applying the associative law gives

$$\begin{aligned} \int_0^t Z_s\, dY_s &= \sum_{i=1}^n \int_0^t \Big(\prod_{m=1, m\ne i}^{n+1} X_s^{k_m} \Big)\, dX_s^{k_i} \\ &\quad + \frac{1}{2} \sum_{i=1}^n \sum_{j=1, j\ne i}^n \int_0^t \Big(\prod_{m=1, m\ne i,j}^{n+1} X_s^{k_m} \Big)\, d\langle X^{k_i}, X^{k_j}\rangle_s. \end{aligned}$$

By definition

$$\int_0^t Y_s\, dZ_s = \int_0^t \left(\prod_{m=1}^{n} X_s^{k_m} \right) dX_s^{k_{n+1}}.$$

To evaluate the third term $\langle Y, Z\rangle_t$ we observe that by our induction hypothesis and Lemma 3.1,

$$\langle Y, Z\rangle_t = \sum_{i=1}^{n} \int_0^t \left(\prod_{m=1, m\neq i}^{n} X_s^{k_m} \right) d\langle X^{k_i}, X^{k_{n+1}}\rangle_s\,.$$

Adding the last three equalities,

$$\begin{aligned} Y_t Z_t - Y_0 Z_0 \quad &= \quad \sum_{i=1}^{n+1} \int_0^t \left(\prod_{m=1, m\neq i}^{n+1} X_s^{k_m} \right) dX_s^{k_i} \\ &\quad + \frac{1}{2} \sum_{i=1}^{n+1} \sum_{j=1, j\neq i}^{n+1} \left(\prod_{m=1, m\neq i,j}^{n+1} X_s^{k_m} \right) d\langle X^{k_i}, X^{k_j}\rangle_s, \end{aligned}$$

which is Itô's formula for $f(x) = x^{k_1} \cdots x^{k_{n+1}}$. This completes the proof. ■

We are now —finally— equipped with enough integration theory to start with the applications. In this lecture I will concentrate efforts on applications to path properties of Brownian motion in one and more dimensions.

Chapter 4

Applications to Brownian motion

4.1 The Dirichlet Problem and exit times

Suppose U is an open, bounded domain in $\mathbb{R}^d$ and let ∂U be its boundary. Suppose that its closure $\overline{U}$ is a homogeneous compact body and its boundary is electrically charged with some continuous function $\phi : \partial U \to \mathbb{R}$ and we are interested in the voltage $u(x)$ at some point $x \in U$. The solution to this problem is (by Kirchhoff's laws) given by the solution to the *Dirichlet problem* with boundary value ϕ. We now show that Brownian motion can be used to solve this problem.

Definition
Let U be an open, bounded domain in $\mathbb{R}^d$ and let ∂U be its boundary. Suppose $\phi : \partial U \to \mathbb{R}$ is a continuous function on its boundary. A function $u : \overline{U} \to \mathbb{R}$, which is twice continuously differentiable on U and continuous on the closure $\overline{U}$ is a *solution to the Dirichlet problem* with boundary value ϕ, if

$$\Delta u(x) = 0 \text{ for all } x \in U$$

and $u(x) = \phi(x)$ for $x \in \partial U$. Here Δ is the Laplace operator $\Delta = \sum_{i=1}^d \partial_{ii}$ and a function with $\Delta u(x) = 0$ for all $x \in U$ is said to be *harmonic on U*.

This problem was posed by Gauss in 1840. In fact Gauss thought he showed that there is always a solution, but his reasoning was wrong and Zaremba in 1909 and Lebesgue in 1913 gave counterexamples. However, if the domain is sufficiently nice there is a solution and in this case the solution can be represented and simulated using Brownian motion. To understand the connection between Brownian motion and harmonic functions, we have a closer look at Itô's formula.

Let $B = \{B(t) = (B^1(t), \ldots, B^d(t)) \; : \; t \geq 0\}$ be a *d-dimensional Brownian motion* started in $x = (x^1, \ldots, x^d)$, i.e. $B^1, \ldots, B^d$ are independent standard Brownian motions started in $x^1, \ldots, x^d$. Recall that

$$\langle B^k \rangle_t = t \text{ and } \langle B^k, B^l \rangle_t = 0 \text{ for all } l \neq k.$$

With this information Itô's formula reads

$$f(B(t)) - f(x) = \int_0^t \sum_{i=1}^d \partial_i f(B(s))\, dB_s^i + \frac{1}{2} \int_0^t \sum_{i=1}^d \partial_{ii} f(B(s))\, ds,$$

for $f : \mathbb{R}^d \to \mathbb{R}$ twice continuously differentiable. Noting that the last integrand is just $\Delta f(B(s))$ we obtain the fundamental relation.

Theorem 4.1 *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is harmonic on U and let $T = \inf\{t \geq 0 : B(t) \notin U\}$. Then $\{f(B(t)) : t \geq 0\}$ is a local martingale on $[0, T)$. More precisely, for all $t < T$,*

$$f(B(t)) = f(x) + \int_0^t \sum_{i=1}^d \partial_i f(B(s))\, dB_s^i.$$

Next we prove one of the highlights of this lecture, the probabilistic representation of the solutions of the Dirichlet problem. This theorem will be the key to several path properties of Brownian motion, which we derive in the sequel.

Theorem 4.2 *Suppose U is an open bounded set and suppose that u is a solution of the Dirichlet problem on U with boundary value ϕ. Define the stopping time*

$$T = \inf\{t \geq 0 : B(t) \notin U\}.$$

Then, for every $x \in U$,

$$u(x) = \mathbb{E}_x\{\phi(B(T))\},$$

where $\mathbb{E}_x$ refers to the expectation with respect to the Brownian motion started in x. In particular, the solution is uniquely determined by the boundary value.

Remarks:

- Our approach is not so useful to show existence of solutions of the Dirichlet problem, because if one defines u as in the theorem, it is difficult to show differentiability. However, if differentiability of this u is assumed, then u solves the Dirichlet problem, which is not hard to show.
- The solution $u(x)$ can be simulated using the formula, by running many independent Brownian motions, starting in $x \in U$ until they hit the boundary of U and letting $u(x)$ be the average of the values of ϕ on the hitting points.
- It is a straightforward consequence of this representation that the solutions u of the Dirichlet problem attain their maximum (and minimum) always on the boundary of the domain.

To prove Theorem 4.2 we first show that the stopping time T is almost surely finite and even has moments of all orders.

Lemma 4.3 *For all $0 < p < \infty$ we have $\sup_{x \in U} \mathbb{E}_x\{T^p\} < \infty$.*

Proof: We start by expressing the moment as a Lebesgue integral and using a change of variable,

$$\mathbf{E}_x\{T^p\} = \int_0^\infty \mathbf{P}_x\{T^p \geq t)\, dt \;\; = \;\; \int_0^\infty ps^{p-1}\, \mathbf{P}_x\{T \geq s\}\, ds.$$

It thus suffices to show that, for some $q < 1$, $\mathbb{P}_x\{T \geq k\} \leq q^k$. To prove this let $K = \sup\{|x-y| : x, y \in U\}$ be the diameter of U. If $x \in U$ and $|B_1 - x| > K$, then $B_1 \notin U$ and hence $T < 1$. Thus

$$\mathbf{P}_x\{T < 1\} \geq \mathbf{P}_x\{|B_1 - x| > K\} = \mathbf{P}_0\{|B_1| > K\} =: \tilde{p} > 0.$$

Letting $q = 1 - \tilde{p}$ we have shown $\mathbf{P}_x\{T \geq 1\} \leq q$, which is the start of an induction argument. Now we can use the (weak) Markov property and the induction hypothesis to infer

$$\mathbf{P}_x\{T \geq k\} \leq \frac{1}{\sqrt{2\pi}^d} \int_U e^{-|x-y|^2/2} \mathbf{P}_y\{T \geq k-1\}\, dy \leq q^{k-1} \mathbf{P}_x\{B_1 \in U\} \leq q^k.$$

This is what we had to show. ■

Proof of Theorem 4.2: As $\Delta u(x) = 0$ for all $x \in U$ we see from our last version of Itô's formula that $u(B(s))$ is a continuous local martingale on the random interval $[0, T)$. Now we look back to Theorem 1.11 and get a time change $\gamma : [0, \infty) \to [0, T)$ such that $X = \{X_t = u(B(\gamma(t))) : t \geq 0\}$ is a bounded martingale with respect to the new filtration defined by $\mathcal{G}(t) := \mathcal{F}(\gamma(t))$. Being a bounded martingale, X converges, by the martingale convergence theorem, almost surely and in L^1 to a limit X_∞ with $X_t = \mathbf{E}_x\{X_\infty | \mathcal{G}(t)\}$. Because T is almost surely finite and u is continuous on the closure of U, we have $X_\infty = u(B(T))$. Hence we infer

$$u(x) = X_0 = \mathbf{E}_x\{X_0\} = \mathbf{E}_x\{X_\infty\} = \mathbf{E}_x\{u(B(T))\}.$$

As $B(T) \in \partial U$ we have that $u(B(T)) = \phi(B(T))$ and we are done. ■

We now apply the theorem to the problem of *recurrence* and *transience* of Brownian motion in various dimensions. Generally speaking, we call a (Markov) process X with values in $\mathbf{R}^d$

- *recurrent*, if for every $x \in \mathbf{R}^d$ there is a (random) sequence $t_n \uparrow \infty$ such that $X(t_n) = x$. We say that x is *visited infinitely often*,
- *neighbourhood recurrent*, if, for every $x \in \mathbf{R}^d$ and $\varepsilon > 0$, the ball $B(x, \varepsilon)$ around x of radius ε is visited infinitely often. Equivalently, every open set is visited infinitely often.
- *transient*, if it converges to infinity almost surely.

The main result of the following discussion will be the following theorem.

Theorem 4.4 *Brownian motion is*

- recurrent *in dimension* $d = 1$,
- neighbourhood recurrent, *but not recurrent, in* $d = 2$,

- transient *in dimension* $d \geq 3$.

In the course of the proof of this we shall derive a couple of further notable facts about Brownian motion. Starting in dimension $d = 1$ we first recover a fact, which already played a big role in the lecture on probability theory.

Theorem 4.5 *Let $a < x < b$ and $T := \inf\{t \geq 0 : B_t \notin (a, b)\}$. Then,*

$$\mathsf{P}_x\{B_T = a\} = \frac{b-x}{b-a} \text{ and } \mathsf{P}_x\{B_T = b\} = \frac{x-a}{b-a}.$$

Proof: Observe that $u(x) = (b - x)/(b - a)$ has $u''(x) = 0$ in $U = (a, b)$ and is continuous on $[a, b]$. It is thus a solution of the Dirichlet problem on U with boundary values $\phi(a) = 1$, $\phi(b) = 0$. From Theorem 4.2 we infer that

$$u(x) = \mathsf{E}_x\{\phi(B_T)\} = \mathsf{P}_x\{B_T = a\},$$

and analogously for the event $\{B_T = b\}$. ■

Corollary 4.6 *Let $T_x := \inf\{t > 0 : B_t = x\}$. Then $\mathsf{P}_y\{T_x < \infty\} = 1$.*

Proof: We may assume that $x = 0$ and, by reflection, $y > 0$. Then, by the last theorem,

$$\mathsf{P}_y\{T_0 < \infty\} \geq \lim_{M\to\infty} \mathsf{P}_y\{T_0 < T_{My}\} = \lim_{M\to\infty} \frac{M-1}{M} = 1.$$

■

By the corollary one dimensional Brownian motion eventually visits every point x. As T_x is a stopping time, by the strong Markov property, $\{B(T_x + t) : t \geq 0\}$ is again a Brownian motion, which visits every point. We wait until the new motion visits a point $y \neq x$, say at time T_y. Then $\{B(T_x + T_y + t) : t \geq 0\}$ is a Brownian motion started in y, which visits x again, and so fourth. With a fixed positive probability it takes at least, say, one time unit before the motion started in y visits x. Because we have infinitely many independent trials for this experiment, there are infinitely many successes by the Borel-Cantelli lemma. This proves that we visit x infinitely often, which means that the process is recurrent.

Let us now move to higher dimension $d \geq 2$, start the motion in x contained in some annulus

$$x \in A := \{x \in \mathsf{R}^d : r \leq |x| \leq R\} \text{ for } 0 < r < R < \infty.$$

What is the probability that the Brownian motion hits the inner ring before it hits the outer ring? In order to copy the proof of the one-dimensional exit problem, we have to find harmonic functions u on the annulus A.

To find them it is first reasonable to assume that u is spherically symmetric, i.e. there is a function $\psi : [r, R] \to \mathsf{R}$ such that $u(x) = \psi(|x|^2)$. We can express derivatives of u in terms of ψ as

$$\partial_i \psi(|x|^2) = \psi'(|x|^2)2x_i \text{ and } \partial_{ii}\psi(|x|^2) = \psi''(|x|^2)4x_i^2 + 2\psi'(|x|^2).$$

Therefore, $\Delta u = 0$ means

$$0 = \sum_{i=1}^{d} \Big(\psi''(|x|^2) 4x_i^2 + 2\psi'(|x|^2) \Big) = 4|x|^2 \psi''(|x|^2) + 2d\psi'(|x|^2).$$

Letting $y = |x|^2 > 0$ we can write this as

$$\psi''(y) = \frac{-d}{2y} \psi'(y).$$

This is solved by every ψ satisfying $\psi'(x) = y^{-d/2}$ and thus $\Delta u = 0$ holds on $\{|x| \neq 0\}$ for

$$u(x) = \begin{cases} |x| & \text{if } d = 1, \\ 2\log|x| & \text{if } d = 2, \\ |x|^{2-d} & \text{if } d \geq 3. \end{cases} \tag{4.1}$$

Now we can use Theorem 4.2. Define stopping times

$$T_r = \inf\{t > 0 \,:\, |B_t| = r\} \text{ for } r > 0.$$

Letting $T = T_r \wedge T_R$ be the first exit time of B from A and abbreviate $u(r)$ for the value of u on $\{|x| = r\}$, we have

$$u(x) = \mathbf{E}_x\{u(B_T)\} = u(r)\mathbf{P}_x\{T_r < T_R\} + u(R)(1 - \mathbf{P}_x\{T_r < T_R\}).$$

This formula can be solved

$$\mathbf{P}_x\{T_r < T_R\} = \frac{u(R) - u(x)}{u(R) - u(r)}$$

and we get an explicit solution for the problem.

Theorem 4.7 *Suppose B is a Brownian motion in dimension $d \geq 2$ started in*

$$x \in A := \{x \in \mathbf{R}^d \,:\, r \leq |x| \leq R\}$$

inside an annulus A with radii $0 < r < R < \infty$. Then, if $d = 2$,

$$\mathbf{P}_x\{T_r < T_R\} = \frac{\log R - \log|x|}{\log R - \log r}.$$

In dimension $d \geq 3$,

$$\mathbf{P}_x\{T_r < T_R\} = \frac{R^{2-d} - |x|^{2-d}}{R^{2-d} - r^{2-d}}.$$

Now look at dimension $d = 2$, fix $r = \varepsilon$ and let $R \uparrow \infty$ in the previous formula. Then we get, for arbitrary $x \notin B(0, \varepsilon)$,

$$\mathbb{P}_x\{T_\varepsilon < \infty\} = \mathbb{P}_x\Big(\bigcup_{R>0} \{T_\varepsilon < T_R\} \Big) = \lim_{R\to\infty} \mathbf{P}_x\{T_\varepsilon < T_R\} = \lim_{R\to\infty} \frac{\log R - \log|x|}{\log R - \log\varepsilon} = 1.$$

Hence the motion eventually hits every small ball around 0. Then, of course, it must hit every small ball eventually and, as before we argue with the strong Markov property to see that it hits every small ball infinitely often. This proves that in $d = 2$ Brownian motion is neighbourhood recurrent.

A compact set $A \subset \mathbb{R}^d$ with the property that

$$\mathbf{P}_x\{B_t \in A \text{ for some } t > 0\} = 0 \text{ for all } x \in \mathbb{R}^d$$

is called a *polar set*. We show that points are polar sets for Brownian motion in $d \geq 2$, which proves that in these dimensions Brownian motion cannot be recurrent.

Theorem 4.8 *For Brownian motion in $d \geq 2$ points are polar sets.*

Proof: It suffices to consider the point 0. Define

$$S_0 = \inf\{t > 0 \,:\, B_t = 0\}.$$

Fix $R > 0$ and let $r \downarrow 0$. First let $x \neq 0$. Then

$$\mathbf{P}_x\{S_0 < T_R\} \leq \lim_{r \to 0} \mathbf{P}_x\{T_r < T_R\} = 0.$$

As this holds for all R and $T_R \to \infty$, by continuity of Brownian motion, we have $\mathbf{P}_x\{S_0 < \infty\} = 0$ for all $x \neq 0$. To extend this to $x = 0$ we observe that the Markov property implies

$$\mathbf{P}_0\{B_t = 0 \text{ for some } t \geq \varepsilon\} = \mathbf{E}_0\Big\{\mathbf{P}_{B_\varepsilon}\{T_0 < \infty\}\Big\} = 0\,,$$

noting that $B_\varepsilon \neq 0$ almost surely. Hence

$$\mathbf{P}_0\{S_0 < \infty\} = \lim_{\varepsilon \downarrow 0} \mathbf{P}_0\{B_t = 0 \text{ for some } t \geq \varepsilon\} = 0.$$

■

Remarks:

- It is clear that for all x and all $t > 0$ we have $\mathbf{P}\{B_t \neq x\} = 1$. Our statement however is that, for all x, $\mathbf{P}\{B_t \neq x \text{ for all } t\} = 1$. This is much harder, because uncountable unions of nullsets usually fail to be nullsets (to see this try to take the quantor for all x inside the probability).
- The proof in the case $d = 2$ also shows that in $d \geq 3$ line segments are polar sets. This does not hold in dimension $d = 2$.

To complete the picture we have to show that Brownian motion is transient in dimensions $d \geq 3$. First observe that the proof of neighbourhood recurrence in dimension $d = 2$ does not work here, because in $d \geq 3$,

$$\mathbf{P}_x\{T_r < \infty\} = \lim_{R \to \infty} \mathbf{P}_x\{T_r < T_R\} = \frac{r^{d-2}}{|x|^{d-2}} < 1$$

for all $|x| > r$. However, we can use this formula to show transience.

Theorem 4.9 *In $d \geq 3$, $\lim_{t\to\infty} B_t = \infty$.*

Proof: Look at the event

$$A_n := \Big\{ |B_t| > \sqrt{n} \text{ for all } t \geq T_n \Big\}$$

and recall that $T_n < \infty$ almost surely by Lemma 4.3. By the strong Markov property we have

$$\mathbf{P}_x(A_n^c) = \mathbf{E}_x\Big\{ \mathbf{P}_{B(T_n)}\{T_{\sqrt{n}} < \infty\} \Big\} = \mathbf{E}_x\Big\{ \lim_{R\to\infty} \mathbf{P}_{B(T_n)}\{T_{\sqrt{n}} < T_R\} \Big\} = \Big(\frac{1}{\sqrt{n}}\Big)^{d-2} \longrightarrow 0.$$

Now let A be the event that infinitely many of the events A_n occur. We have

$$\mathbf{P}_x(A) = \mathbf{P}_x\Big(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \Big) = \limsup_{n\to\infty} \mathbf{P}_x\Big(\bigcup_{k=n}^{\infty} A_k \Big) \geq \limsup_{n\to\infty} \mathbf{P}_x(A_n) = 1\,,$$

hence A holds almost surely, which means that for infinitely many n the path eventually does not return inside the ball of radius $\sqrt{n}$, so it must go to infinity almost surely. ∎

To summarise what we have learned in this chapter so far, we recall that the recurrence and transience behaviour of Brownian motion depend only on the behaviour of the harmonic functions u on the annulus in $\mathbb{R}^d$, which are defined in (4.1). The proof of neighbourhood recurrence was based on

$$\lim_{x\to\infty} u(x) = \infty \text{ if } d = 2,$$

whereas the proof of polarity of the points used

$$\lim_{x\to 0} |u(x)| = \infty \text{ in the case } d \geq 2,$$

and, finally, the proof of transience was based on

$$\lim_{x\to\infty} u(x) = 0 \text{ if } d \geq 3.$$

Exit problems from domains are closely linked to the behaviour of harmonic functions (and the Laplace operator) on the domain. We conclude this chapter with one more problem from this realm.

Suppose we start a d-dimensional Brownian motion in some point x inside an open bounded domain U. Let

$$T = \inf\{t \geq 0 \,:\, B(t) \notin U\}.$$

We ask for the distribution of the point $B(T)$ where it first leaves U. This distribution on the boundary ∂U is usually called the *harmonic measure on the domain* U. Of course it depends on the starting point x, but by a famous theorem of Harnack all these measures are absolutely continuous with respect to each other.

We will concentrate on the case of exit distributions from the unit ball

$$U = \{x \in \mathbb{R}^d \,:\, |x| < 1\}.$$

If $x = 0$ the distribution of $B(T)$ is (by symmetry) the uniform distribution π, but if x is another point it is an interesting problem to determine this distribution in terms of a probability density. The solution is again closely related to the Dirichlet problem.

Theorem 4.10 (Poisson's formula) *Suppose that $A \subset \partial U$ is a Borel subset of the unit sphere $\partial U \subset \mathbb{R}^d$ for $d \geq 2$. Then, for all $x \in U$,*

$$\mathbf{P}_x\{B(T) \in A\} = \int_A \frac{1-|x|^2}{|x-y|^d}\, d\pi(y),$$

where π denotes the uniform distribution on the unit sphere.

Remark: The density appearing in the theorem is usually called the *Poisson kernel* and appears frequently in potential theory.

Proof: To prove the theorem we indeed show that for every bounded measurable $f : \mathbb{R}^d \to \mathbb{R}$ we have

$$\mathbf{E}_x\{f(B(T))\} = \int_{\partial U} \frac{1-|x|^2}{|x-y|^d} f(y)\, d\pi(y), \tag{4.2}$$

which on the one hand implies the formula by choosing indicator functions, on the other hand, by the monotone class theorem, it suffices to show this for C^∞-functions. To prove (4.2) we recall Theorem 4.2, which tell us that we just have to show that the right hand side as a function in $x \in U$ defines a solution of the Dirichlet problem on U with boundary value f.

To check this, one first checks that $\frac{1-|x|^2}{|x-y|^d}$ is harmonic on U, which is just a calculation, and then argues that it is allowed to differentiate twice under the integral sign. We omit the slightly boring details, referring readers to Durrett p.164. To check the boundary condition first look at the case $f \equiv 1$. Then we have to show that, for all $x \in U$,

$$I(x) := \int_{\partial U} \frac{1-|x|^2}{|x-y|^d}\, \pi(dy) \equiv 1.$$

This can be computed as well, but we argue mathematically. Observe that $I(0) = 1$, I is invariant under rotation and $\Delta I = 0$ on U, by the first part. Now let $x \in U$ with $|x| = r < 1$ and let $\tau := \inf\{t : |B_t| > r\}$. By Theorem 4.2

$$I(0) = \mathbf{E}_0\{I(B_\tau)\} = I(x)\,,$$

using rotation invariance in the second step. Hence $I \equiv 1$. Now we show that the right hand side in the theorem can be extended continuously to all points $y \in \partial U$ by $f(y)$. We write D_0 for ∂U with a δ-neighbourhood $U(y,\delta)$ removed and $D_1 = \partial U \setminus D_0$. We have, using that $I \equiv 1$, for all $x \in U(y, \delta/2) \cap U$,

$$\begin{aligned}
&\Big| f(y) - \int_{\partial U} \frac{1-|x|^2}{|x-z|^d} f(z)\, d\pi(z) \Big| \\
&= \Big| \int_{\partial U} \frac{1-|x|^2}{|x-z|^d} (f(y) - f(z))\, d\pi(z) \Big| \\
&\leq 2\|f\|_\infty \int_{D_0} \frac{1-|x|^2}{|x-z|^d}\, d\pi(z) + \sup_{z \in D_1} |f(y) - f(z)|.
\end{aligned}$$

For fixed $\delta > 0$ the first term goes to 0 as $x \to y$ by dominated convergence, whereas the second can be made arbitrarily small by choice of δ. ■

4.2 The Poisson problem and occupation times

In this chapter we will attack the problem, how much time a d-dimensional Brownian motion spends in a domain U before it leaves the domain. These *occupation times* of Brownian motion are —via stochastic integration— linked to another classical problem of partial differential equations, the *Poisson problem.*

Definition Let U be an open bounded domain and $u : \overline{U} \to \mathbb{R}$ be a continuous function, which is twice continuously differentiable on U. Let $g : U \to \mathbb{R}$ be continuous. Then u is said to be the *solution of Poisson's problem for* g if $u(x) = 0$ for all $x \in \partial U$ and

$$\frac{\Delta}{2} u(x) = -g(x) \text{ for all } x \in U.$$

Theorem 4.11 *Suppose g is bounded and u a bounded solution of Poisson's problem for g. Then this solution has the form*

$$u(x) = \mathbf{E}_x\Big\{ \int_0^T g(B_t)\, dt \Big\} \textit{ for } x \in U\,,$$

where $T := \inf\{t > 0 \,:\, B(t) \notin U\}$. In particular, the solution, if it exists, is always uniquely determined.

Remark: If $g \equiv 1$, then $u(x) = \mathbf{E}_x\{T\}$.

This relates the solutions u of Poisson's problem to the time spent by Brownian motion in U before leaving the domain. To prove the relation we proceed similarly as in the case of the Dirichlet problem, we first use Itô's formula to find a local martingale.

Theorem 4.12 *Let U be a bounded open domain and $T := \inf\{t > 0 \,:\, B(t) \notin U\}$. If u is a solution to Poisson's problem for g, then $M = \{M_t \,:\, t \geq 0\}$ with*

$$M_t := u(B(t)) + \int_0^t g(B(s))\, ds$$

defines a local martingale on $[0, T)$.

Proof: Applying Poisson's equation and Itô's formula gives, for all $t < T$,

$$\begin{aligned} u(B(t)) + \int_0^t g(B(s))\, ds &= u(B(t)) - \frac{1}{2} \int_0^t \Delta u(B(s))\, ds \\ &= u(B(0)) + \int_0^t \sum_{i=1}^d \partial_i u(B(s))\, dB^i(s), \end{aligned}$$

which is a local martingale on $[0, T)$. ■

Proof of Theorem 4.11: M is a local martingale on $[0, T)$ and we let $\gamma : [0, \infty) \to [0, T)$ be the time change such that $\{M(\gamma(t)) : t \geq 0\}$ is a martingale. As u and g are bounded,

$$\sup_{t \geq 0} |M(\gamma(t))| \leq \|u\|_\infty + T\|g\|_\infty.$$

The right hand side is L^2-integrable by Lemma 4.3 and hence we have that $\{M(\gamma(t)) : t \geq 0\}$ is a uniformly integrable martingale. The martingale convergence theorem states that $\lim_{t\to\infty} M(\gamma(t)) =: M_\infty$ exists almost surely and $\mathbb{E}_x\{M_0\} = \mathbb{E}_x\{M_\infty\}$. We can now use continuity of u and g to get

$$u(x) = \mathbb{E}_x\{M_0\} = \mathbb{E}_x\{M_\infty\} = \mathbb{E}_x\Big\{ \lim_{t\uparrow T} u(B_t) + \int_0^t g(B_s)\, ds \Big\} = \mathbb{E}_x\Big\{ \int_0^T g(B_s)\, ds \Big\},$$

as stated in the theorem. ■

We now address the following question: given a bounded open domain $U \subset \mathbb{R}^d$, does Brownian motion spend an infinite or a finite amount of time in U? It is not surprising that the answer depends on the dimension, more interestingly, it depends neither on U nor on the starting point.

Theorem 4.13 *Let $U \subset \mathbb{R}^d$ be a bounded open domain and $x \in \mathbb{R}^d$ arbitrary. Then, if $d \leq 2$,*

$$\int_0^\infty 1_U(B_t)\, dt = \infty \qquad \mathbb{P}_x\text{-almost surely,}$$

and, if $d \geq 3$,

$$\mathbb{E}_x \int_0^\infty 1_U(B_t)\, dt < \infty.$$

Proof: As U is contained in a ball, and contains a ball, it suffices to show this for balls. By shifting, we can even restrict to balls $U = B(0,r)$ centred in the origin. Let us start with the first claim. We let $d \leq 2$ and let $G = B(0, 2r)$. Let $T_0 = 0$ and, for all $k \geq 1$, let

$$S_k = \inf\{t > T_{k-1} : B_t \in U\} \text{ and } T_k = \inf\{t > S_k : B_t \notin G\}.$$

From the strong Markov property we infer, for $k \geq 1$,

$$\mathbb{P}_x\Big\{ \int_{S_k}^{T_k} 1_U(B_t)\, dt \geq s \,\Big|\, \mathcal{F}(S_k) \Big\} = \mathbb{P}_{B(S_k)}\Big\{ \int_0^{T_1} 1_U(B_t)\, dt \geq s \Big\}$$

$$= \mathbb{E}_x\Big\{ \mathbb{P}_{B(S_k)}\Big\{ \int_0^{T_1} 1_U(B_t)\, dt \geq s \Big\} = \mathbb{P}_x\Big\{ \int_{S_k}^{T_k} 1_U(B_t)\, dt \geq s \Big\},$$

by rotation invariance. The second expression does not depend on k, so that the random variables

$$\int_{S_k}^{T_k} 1_U(B_t)\, dt$$

are independent and identically distributed. As they are not identically zero, but nonnegative, they have positive expected value and, by the strong law of large numbers we infer,

$$\int_0^\infty 1_U(B_t)\, dt = \lim_{n\to\infty} \sum_{k=1}^n \int_{S_k}^{T_k} 1_U(B_t)\, dt = \infty,$$

which proves the first claim. For the second claim, first let f be nonnegative and measurable. Fubini's Theorem implies

$$\begin{aligned}\mathsf{E}_x \int_0^\infty f(B_t)\,dt &= \int_0^\infty \mathsf{E}_x f(B_t)\,dt = \int_0^\infty \int p_t(x,y) f(y)\,dy\,dt \\ &= \int \int_0^\infty p_t(x,y)\,dt\, f(y)\,dy,\end{aligned}$$

where $p_t(x,y) = (2\pi t)^{-d/2}\exp(-|x-y|^2/2t)$ is the transition density of Brownian motion. Note that for large t we have $p_t(x,y) \sim (2\pi t)^{-d/2}$, hence $\int_0^\infty p_t(x,y)\,dt = \infty$ if $d \le 2$. In $d \ge 3$ however we can define, for $x \neq y$,

$$G(x,y) := \int_0^\infty p_t(x,y)\,dt < \infty\,,$$

a quantity called *the potential kernel.* It can be calculated explicitly, changing variables $s = |x-y|^2/2t$,

$$\begin{aligned}G(x,y) &= \int_0^\infty \frac{1}{(2\pi t)^{-d/2}} e^{-|x-y|^2/2t}\,dt = \int_\infty^0 \Big(\frac{s}{\pi|x-y|^2}\Big)^{d/2} e^{-s}\Big(-\frac{|x-y|^2}{2s^2}\Big)\,ds \\ &= \frac{|x-y|^{2-d}}{2\pi^{d/2}} \int_0^\infty s^{(d/2)-2} e^{-s}\,ds = \frac{\Gamma(d/2-1)}{2\pi^{d/2}}|x-y|^{2-d},\end{aligned}$$

where $\Gamma(x) = \int_0^\infty s^{x-1}e^{-s}\,ds$ is the gamma function. We abbreviate the constant in the last term by $c(d)$. To summarise what we have done so far,

$$\mathsf{E}_x \int_0^\infty f(B_t)\,dt = \int G(x,y) f(y)\,dy = c(d) \int \frac{f(y)}{|x-y|^{d-2}}\,dy. \tag{4.3}$$

To finish the proof let $f = 1_{B(0,r)}$. If $x = 0$ we may change to polar coordinates, writing $C(d)$ for a constant depending only on d,

$$\mathsf{E}_0 \int_0^\infty 1_{B(0,r)}(B_s)\,ds = \int_{B(0,r)} G(0,y)\,dy = C(d)\int_0^r s^{d-1}s^{2-d}\,ds = (C(d)/2)r^2 < \infty.$$

For start in an arbitrary $x \neq 0$ we look at a Brownian motion started in 0 and a stopping time T, which is the first hitting time of the sphere $\partial B(0,|x|)$. Using spherical symmetry and the strong Markov property we obtain

$$\mathsf{E}_x \int_0^\infty 1_{B(0,r)}(B_s)\,ds = \mathsf{E}_0 \int_T^\infty 1_{B(0,r)}(B_s)\,ds \le \mathsf{E}_0 \int_0^\infty 1_{B(0,r)}(B_s)\,ds < \infty.$$

■

We shall now study the *expected occupation measures* of a Brownian motion in a Borel subset A of some bounded open domain U. These are defined as

$$\mathsf{E}_x \int_0^T 1_A(s)\,ds \text{ for } A \subset U \text{ Borel},$$

in other words the expected time Brownian motion spends in A before leaving the domain U.

Theorem 4.14 *Let $d \geq 3$ and $U \subset \mathbf{R}^d$ a bounded domain, let T be the first exit time from the domain. Define the potential kernel $G(x,y) = c(d)|x-y|^{2-d}$ for $c(d) = \frac{\Gamma(d/2-1)}{2\pi^{d/2}}$, for $x \in U$ recall the definition of the harmonic measure*

$$\mu(x, dz) = \mathbf{E}_x\{B_T \in dz\}$$

and define the Green's function *of the domain U as*

$$G_U(x,y) = G(x,y) - \int_{\partial U} G(z,y)\, \mu(x,dz) \text{ for } x,y \in U.$$

Then we have, for all $x \in U$ and $A \subset U$ Borel,

$$\mathbf{E}_x\Big\{ \int_0^T 1_A(B_s)\, ds \Big\} = \int_A G_U(x,y)\, dy.$$

Remarks:

- Probabilistically the Green's function $G_U(x,y)$ is the density of the expected occupation measure of a Brownian motion started in x and stopped upon leaving U.
- Note that the theorem also says that the bounded solution of the Poisson problem for g is $\int G_U(x,y)g(y)\, dy$. This is the P.D.E. point of view of the Green's function.
- Letting $g(y)\, dy$ converge to the Dirac measure δ_y we get the physical interpretation of the Green's function $G_U(x,y)$ as the electrostatic potential of a unit mass at y with ∂U grounded.
- Although in $d = 1,2$ it is not possible to put $G(x,y) = \int_0^\infty p_t(x,y)\, dt$, it is also possible to define a potential kernel G such that the theorem holds true. As in $d = 3$ these potential kernels are just constant multiples of $u(x-y)$ for the harmonic functions u on the punctured disk, see (4.1).

Proof: Suppose that g is a nonnegative, bounded function with compact support. Then, by Theorem 4.13,

$$w(x) := \mathbf{E}_x \int_0^\infty g(B_s)\, ds < \infty.$$

By the strong Markov property,

$$w(x) = \mathbf{E}_x \int_0^T g(B_s)\, ds + \mathbf{E}_x w(B_T).$$

We already know that

$$w(x) \quad = \quad \int_0^\infty \mathbf{E}_x g(B_s)\, ds = \int_0^\infty \int p_s(x,y) g(y)\, dy\, ds = \int G(x,y)\, g(y)\, dy.$$

Inserting this gives

$$\mathbf{E}_x\Big\{ \int_0^T g(B_s)\, ds \Big\} = \int G(x,y)\, g(y)\, dy - \mathbf{E}_x \int G(B_T,y)\, g(y)\, dy = \int G_U(x,y) g(y)\, dy,$$

which is the required formula. ■

In the case $U = B(0,1)$ this can be calculated explicitly, using Theorem 4.11. We have

$$\begin{aligned} G_U(x,y) &= G(x,y) - \int_{\partial U} \frac{1-|x|^2}{|x-z|^d}\, G(z,y)\, \pi(dz) \\ &= \frac{\Gamma(d/2-1)}{2\pi^{d/2}} \int_{\partial U} \frac{1-|x|^2}{|x-z|^d} \Big(|x-y|^{2-d} - |z-y|^{2-d} \Big) \pi(dz). \end{aligned}$$

The latter integral can be evaluated explicitly.

Theorem 4.15 *Let $d \geq 3$ and $U \subset \mathbb{R}^d$ the unit disc, T be the first exit time from U. Then,*

$$\int_{\partial U} \frac{1-|x|^2}{|x-z|^d}\, G(z,y)\, \pi(dz) = c(d) \frac{|y|^{d-2}}{|\, x|y|^2 - y\,|^{d-2}} \text{ for all } x, y \in U\ .$$

Proof: Of course, we do not calculate much to evaluate the integral and find G_U. We already know from the proof of Poisson's formula and the definition of G_U that, for every $y \in U$,

$$u(x) := G(x,y) - G_U(x,y) = \int_{\partial U} \frac{1-|x|^2}{|x-z|^d} G(z,y) \pi(dz)$$

can be extended continuously to the boundary of U to give the solution for the Dirichlet problem with boundary function $\phi(x) = G(x,y)$. By uniqueness it suffices to check that also

$$v(x) := c(d)\, \frac{|y|^{d-2}}{|\, x|y|^2 - y\,|^{d-2}}$$

has this property, so that $v = u$ and we are done. To see this note that

$$v(x) = |y|^{2-d} G(x, y/|y|^2).$$

Because $G(x,y)$ is just a constant multiple of the function $u(x-y)$ defined in (4.1) and u was harmonic on the punctured disk, we infer from $y/|y|^2 \notin U$ that v is harmonic on U. Clearly, v is continuous on $\overline{U}$. To determine the value on the boundary observe that, if $|x| = 1$,

$$|\, x|y| - y/|y|\,|^2 = |x|^2|y|^2 - 2x \cdot y + 1 = |y|^2 - 2x \cdot y + |x|^2 = |x-y|^2.$$

Therefore,

$$\begin{aligned} G(x,y) - |y|^{2-d} G(x, y/|y|^2) &= \frac{c(d)}{|x-y|^{d-2}} - \frac{c(d)}{|y|^{d-2}} \Big| x - \frac{y}{|y|^2} \Big|^{2-d} \\ &= \frac{c(d)}{|x-y|^{d-2}} - \frac{c(d)}{|\, x|y| - y/|y|\,|^{d-2}} = 0. \end{aligned}$$

This concludes the proof of the equality $u = v$. ■

Example: Suppose we start a three dimensional Brownian motion at the origin. Fix a radius $0 < r < 1$. In which of the annuli

$$A[a] = \{x \in \mathbb{R}^3 : a - r \le |x| \le a\} \text{ for } a \in [r, 1]$$

is the expected occupation time maximal?

We can do the calculation in polar coordinates, letting C denote $c(d)$ times the surface of the unit sphere in $\mathbb{R}^3$. Then

$$\begin{aligned} \mathbb{E}_0\Big\{ \int_0^T 1_{A[a]}(B_s)\, ds \Big\} &= C \int_{a-r}^{r} \Big(\frac{1}{r} - 1\Big)\, r^2\, dr \\ &= C\Big(\frac{a^2}{2} - \frac{a^3}{3} - \frac{(a-r)^2}{2} + \frac{(a-r)^3}{3}\Big) \\ &= C\Big(r(a - a^2) + r^2(a - 1/2) - r^3/3\Big). \end{aligned}$$

We note that this is maximal if $r(1-2a) + r^2 = 0$ i.e. exactly if $a = (1+r)/2$, if the annulus is in the middle. Note that the expected occupation time in the inner ball $a = r$ and the outer annulus $a = 1$ are the same.

4.3 The Feynman-Kac formula

We will study one more partial differential equation by a probabilistic representation. This time we look at a time inhomogeneous setting and see that certain parabolic equations can also be treated with our method. A continuous function $u : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}$, which is twice continuously differentiable, is said to satisfy the *heat equation with heat dissipation rate* $c : (0, \infty) \times \mathbb{R}^d \to \mathbb{R}$ if we have

- u is continuous at each point of $\{0\} \times \mathbb{R}^d$ and $u(0, x) = f(x)$,
- $\partial_t u(t, x) = \frac{1}{2}\Delta u(t, x) - c(t, x)u(t, x)$ on $(0, \infty) \times \mathbb{R}^d$.

If $c(t, x) \ge 0$ then $u(t, x)$ describes the temperature at time t at x for a heat flow with cooling. Here the initial temperature distribution is given by f and $c(t, x)$ describes the heat dissipation rate at time t in x. We start our machinery, as usual, by finding a local martingale.

Theorem 4.16 *If u is a solution of the heat equation with dissipation rate c, then $M = \{M_s : s \ge 0\}$ with*

$$M_s := u(t - s, B_s) \exp\Big(- \int_0^s c(t - r, B_r)\, dr\Big)$$

is a local martingale on $[0, t)$.

Proof: Abbreviate $c^t_s = -\int_0^s c(t - r, B_r)\, dr$. We apply Itô's formula with the semimartingales $X^0_s = t - s$, $X^i_s = B^i_s$ for $1 \le i \le d$ and $X^{d+1}_s = c^t_s$.

$$u(t - s, B_s) \exp(c^t_s) - u(t, B_0)$$

$$= \int_0^s -\partial_t u(t-r,B_r)\exp(c_r^t)\,dr + \sum_{j=1}^d \int_0^s \exp(c_r^t)\partial_j u(t-r,B_r)\,dB_r^j$$
$$+ \int_0^s u(t-r,B_r)\exp(c_r^t)\,dc_r^t + \frac{1}{2}\int_0^s \Delta u(t-r,B_r)\exp(c_r^t)\,dr,$$

since we have

$$\langle X^i, X^j\rangle_t = \begin{cases} t & \text{if } 1 \le i = j \le d \\ 0 & \text{otherwise.} \end{cases}$$

Using $dc_r^t = -c(t-r,B_r)\,dr$ and rearranging, the right hand side is

$$= \int_0^s \Big(-\partial_t u - cu + \frac{\Delta}{2}u\Big)(t-r,B_r)\exp(c_r^t)\,dr + \sum_{j=1}^d \int_0^s \exp(c_r^t)\partial_j u(t-r,B_r)\,dB_r^j.$$

This proves the claim, because $-\partial_t u - cu + \frac{\Delta}{2}u = 0$ and the second term is a local martingale. ■

The resulting representation theorem for the solutions of our heat equation with dissipation term is called the Feynman-Kac formula.

Theorem 4.17 (Feynman-Kac formula) *Suppose that the dissipation rate c is bounded and u is a solution of the heat equation with dissipation rate c, which is bounded on every set $[0,t]\times\mathbb{R}^d$. Then u is uniquely determined and satisfies*

$$u(t,x) = \mathbf{E}_x\Big\{ f(B_t)\exp\Big(-\int_0^t c(t-r,B_r)\,dr\Big)\Big\}.$$

Proof: Under our assumptions on c and u, M is a bounded martingale on $[0,t)$ and $M_t = \lim_{s\uparrow t} M_s = f(B_t)\exp(c_t^t)$. Since M is uniformly integrable we infer that

$$\mathbf{E}_x\Big\{ f(B_t)\exp\Big(-\int_0^t c(t-r,B_r)\,dr\Big)\Big\} = \mathbf{E}_x\{M_t\} = \mathbf{E}_x\{M_0\} = u(t,x).$$

■

Example: Brownian motion in a soft potential. Imagine that the bounded function $c : \mathbf{R}^d \to [0,\infty)$ defines a potential landscape, so that it is hard for a particle to go through the hills of c and easier to go through the valleys. To model this, suppose that a Brownian path, which has gone through a certain amount of hills will be killed with a probability, which increases with this amount. For example, for a path $\{B(t) : t \ge 0\}$ we suppose that the probability of survival up to time T is

$$\exp\Big(-\int_0^T c(B_r)\,dr\Big).$$

It is possible to construct such a process $\{X(t) : t \ge 0\}$ with values in $\mathbf{R}^d \cup \{\dagger\}$, called a *killed Brownian motion.* If $u_1 : [0,\infty)\times\mathbf{R}^d \to \mathbf{R}$ is a solution of our problem with dissipation rate c and initial value $f \equiv 1$, which is bounded on the set $[0,T]\times\mathbf{R}^d$, then the probability that a path $\{X(t) : t \ge 0\}$ started in x survives up to time T is

$$u_1(T,x) := \mathbf{E}_x\Big\{ \exp\Big(-\int_0^T c(B_r)\,dr\Big)\Big\}.$$

Suppose now f is a bounded function and we are interested in the expected value of $f(X(T))$ for the killed Brownian motion X conditioned on survival up to time T. First note that we do not expect the result to be the same as with ordinary Brownian motion, as killed Brownian motion conditioned on survival up to time T is intuitively more likely to go through the valleys of c than through the hills. Suppose that $u_2 : [0,\infty) \times \mathbf{R}^d \to \mathbf{R}$ is a solution of our problem with initial value f, which is bounded on $[0,T] \times \mathbf{R}^d$, then

$$\mathbf{E}_x\{f(X(T)) \,|\, X(T) \neq \dagger\} = \frac{\mathbf{E}_x\{f(B_T) \exp\big(-\int_0^T c(B_r)\,dr\big)\}}{\mathbf{E}_x\{\exp\big(-\int_0^T c(B_r)\,dr\big)\}} = \frac{u_2(T,x)}{u_1(T,x)}.$$

The process X conditioned on survival up to time T, i.e. on $\{X(t) \neq \dagger\}$, is called *Brownian motion in the soft potential* c and, if we denote it by $Y = \{Y_t : t \in [0,T]\}$, its marginal distributions at the endpoint are given by the Feynman-Kac formula as

$$\mathbf{E}\{f(Y(T))\} = \frac{u_2(T,x)}{u_1(T,x)}.$$

This method of defining stochastic processes is very useful to construct processes with prescribed properties and the Feynman-Kac formula plays a key role.

4.4 Change of measure and Girsanov's formula

In the last example we have defined a new stochastic process $\{Y(s) : s \in [0,T]\}$, Brownian motion in the potential c, on the space $C[0,T]$ of continuous paths $f : [0,T] \to \mathbf{R}^d$ by means of a probability density

$$\alpha(f) = \frac{\exp\big(-\int_0^T c(f(r))\,dr\big)}{\mathbf{E}_x\{\exp\big(-\int_0^T c(B(r))\,dr\big)\}}$$

with respect to the Wiener measure on $C[0,T]$. Intuitively, it is clear that for most potential landscapes c the new process is not a martingale any more. In this section we shall see that it is always a semimartingale.

To discuss general criteria whether stochastic processes defined by changing the measure on path space are martingales or semimartingales, we fix some notation. We let $C := C[0,T]$ or $C[0,\infty)$ the space of continuous paths equipped with the Borel σ-field $\mathcal{C}$ and the right-continuous filtration $\{\mathcal{F}_t\}$ such that

$$\mathcal{F}_t := \bigcap_{\varepsilon>0} \sigma\Big\{X_s : s \leq t+\varepsilon\Big\} \text{ where } X_s(f) = f(s)$$

is the coordinate map. We denote the restrictions of P and Q to $\mathcal{F}_t$ by P_t and Q_t. Two measures P and Q on C are called *locally equivalent* if for each t the measures P_t and Q_t have the same zerosets. In this case one can define the density

$$\alpha_t = \frac{dQ_t}{dP_t} : C \to [0,\infty),$$

using the Radon-Nikodym Theorem. Now $\alpha = \{\alpha_t : t \geq 0\}$ is a stochastic process on $(C, \mathcal{C}, P)$, which is adapted to our filtration. We generally assume that this process has almost surely continuous paths.

Remark: The last assumption is automatically fulfilled in many cases, see for example Durrett (6.1) in Chapter 3.

EXAMPLE: Let $P = \mathbf{P}_x$ and define $dQ = \alpha\, dP$ as in the motivating example of Brownian motion in a soft potential c. We abbreviate

$$Z := \mathbf{E}_x\Big\{ \exp\Big(- \int_0^T c(B_r)\, dr\Big)\Big\}.$$

Then, for every $A \in \mathcal{F}_t$,

$$\begin{aligned} Q_t(A) &= Q(A) = \frac{1}{Z}\mathbf{E}_x\Big\{1_A(B) \exp\Big(- \int_0^T c(B_r)\, dr\Big)\Big\} \\ &= \frac{1}{Z}\mathbf{E}_x\Big\{1_A(B) \exp\Big(- \int_0^t c(B_r)\, dr\Big)\mathbf{E}_{B(t)}\Big\{ \exp\Big(- \int_0^{T-t} c(\tilde{B}_r)\, dr\Big)\Big\}\Big\}, \end{aligned}$$

where we have used the Markov property in the last step. This means that

$$\alpha_t = \frac{1}{Z} \exp\Big(- \int_0^t c(B_r)\, dr\Big)\mathbf{E}_{B(t)}\Big\{ \exp\Big(- \int_0^{T-t} c(\tilde{B}_r)\, dr\Big)\Big\}.$$

α is continuous and one can check that it is a martingale, but this also follows from the general theory, as we shall see below.

If Y is an adapted process on the measurable space $(C, \mathcal{C})$ we say that Y *is a martingale[P]* if Y is a martingale on the probability space $(C, \mathcal{C}, P)$.

Lemma 4.18 *If P and Q are locally equivalent, then an adapted process $Y = \{Y_t : t \geq 0\}$ is a martingale$[Q]$ if and only if $\alpha Y = \{\alpha_t Y_t : t \geq 0\}$ is a martingale$[P]$.*

Proof: Suppose Y is a martingale$[Q]$, $s < t$ and $A \in \mathcal{F}_s$. Then

$$\begin{aligned} \int_A \alpha_t Y_t\, dP &= \int_A \alpha_t Y_t\, dP_t = \int_A Y_t\, dQ_t = \int_A Y_t\, dQ \\ &= \int_a Y_s\, dQ = \int_A Y_s\, dQ_s = \int_A \alpha_s Y_s\, dP_s = \int_A \alpha_s Y_s\, dP. \end{aligned}$$

This proves that $\mathbf{E}\{\alpha_t Y_t \,|\mathcal{F}_s\} = \alpha_s Y_s$ and hence αY is a martingale$[P]$. The converse follows by exchanging the roles of P and Q using that $1/\alpha_t = dP_t/dQ_t$. ■

Because 1 is a martingale$[Q]$ we infer that α is always a martingale$[P]$.

Lemma 4.19 *A process $Y = \{Y_t : t \geq 0\}$ is a local martingale$[Q]$ if and only if $\alpha Y = \{\alpha_t Y_t : t \geq 0\}$ is a local martingale$[P]$.*

Proof: If Y is a local martingale$[Q]$ and T_n a reducing sequence of stopping times then, by the last lemma, $\alpha . Y_{.\wedge T_n}$ is a martingale$[P]$. By the optional stopping theorem we infer that also $\alpha_{.\wedge T_n} Y_{.\wedge T_n}$ defines a martingale$[P]$ and hence the statement follows. Again the converse is analogous. ■

The main result of this section implies, for example, that our Brownian motion in a soft potential is a semimartingale.

Theorem 4.20 (Girsanov's formula) *If X is a local martingale$[P]$ and*

$$A_t := \int_0^t \alpha_s^{-1}\, d\langle \alpha, X\rangle_s \,,$$

then $X - A$ is a local martingale$[Q]$ and, in particular, X is a semimartingale with respect to Q.

Corollary 4.21 *A process is a semimartingale$[P]$ if and only if it is a semimartingale$[Q]$, so that the notion of a semimartingale is not affected by a locally equivalent change of measure.*

For the proof of Girsanov's formula we need the following lemma.

Lemma 4.22 *Let $S_n := \inf\{t : \alpha_t \le \frac{1}{n}\}$ and $S = \lim_{n\to\infty} S_n$. Then $\alpha_t = 0$ almost surely on $\{S \le t\}$.*

Proof: By the Optional Stopping Theorem we have $\mathbf{E}\{\alpha_{t\wedge S_n}\} = \mathbf{E}\{\alpha_t\}$ and, as $\alpha_{t\wedge S_n} = \alpha_t$ on $S_n > t$ we have

$$\mathbf{E}\Big\{\alpha_t 1_{\{S_n \le t\}}\Big\} = \mathbf{E}\Big\{\alpha_{S_n} 1_{\{S_n \le t\}}\Big\}.$$

Since α is continuous, nonnegative and $S_n \le S$ we infer

$$\mathbf{E}\Big\{\alpha_t 1_{\{S \le t\}}\Big\} \le \mathbf{E}\Big\{\alpha_t 1_{\{S_n \le t\}}\Big\} = \mathbf{E}\Big\{\alpha_{S_n} 1_{\{S_n \le t\}}\Big\} \le \frac{1}{n}.$$

Letting $n \to \infty$ we see that $\alpha_t = 0$ almost surely on $\{S \le t\}$. ∎

Proof of the Girsanov formula: We first check that the process A defined in the theorem by means of an integral has finite values. If $t \le S_n$ then,

$$\Big| \int_0^t \alpha_s^{-1}\, d\langle \alpha, X\rangle_s \Big| \le \int_0^t \frac{dV_s}{\alpha_s} \le n \int_0^t dV_s \,,$$

where V_s is the variation of $\langle \alpha, X\rangle$ on $[0, s]$. By the Kunita-Watanabe inequality, the last integral is bounded by

$$(\langle \alpha\rangle_t)^{1/2} (\langle X\rangle_t)^{1/2} < \infty.$$

Hence A is well defined if $t < S$ and, on the other hand, by the lemma,

$$Q_t\{S \le t\} = \mathbf{E}\{\alpha_t 1_{\{S \le t\}}\} = 0.$$

Hence, $\{S \le t\}$ is a Q_t nullset and $S > t$ Q-almost surely (and P-almost surely) and we are done.

To check that the process A satisfies the requirements of Girsanov's formula, we apply the integration by parts rule with respect to P.

$$\begin{aligned}
\alpha_t(X_t - A_t) - \alpha_0 X_0 &= \int_0^t (X_s - A_s)\, d\alpha_s + \int_0^t \alpha_s\, dX_s - \int_0^t \alpha_s\, dA_s + \langle \alpha, X\rangle_t. \\
&= \int_0^t (X_s - A_s)\, d\alpha_s + \int_0^t \alpha_s\, dX_s.
\end{aligned}$$

As the expression on the right defines a local martingale[P] we are done. ■

As the notion of a semimartingale is not affected by locally equivalent change of measure, one might ask whether the stochastic integrals are independent of this choice, too. The next theorem shows that this is indeed the case.

Theorem 4.23 *If $H \in \Pi$ is a locally bounded and predictable process, then $(H \cdot X)$ is the same random variable on C, no matter whether P or Q where chosen for the definition. Also, the definition of the variance process $\langle X \rangle$ does not depend on the choice of a locally equivalent measure.*

Proof: Recall that if $\Delta_n = \{0 = T_0^n < \cdots < t_{k_n}^n = t\}$ is a sequence of partitions with mesh going to 0, then

$$\sum_{i=1}^{n} (X(t_{i+1}^n) - X(t_i^n))^2 \to \langle X \rangle_t$$

in probability[P] for any semimartingale. The second statement follows, because convergence in probability is not affected by changing to an equivalent measure. Now we look at the integrals and let $X = M + A$ be the semimartingale decomposition of X with respect to P any $X = N + B$ be the semimartingale decomposition with respect to Q. We already know from the initial observation that $\langle M \rangle_t = \langle N \rangle_t$. If the integrand H is simple, then the definition of the integral does not refer to the measure and the values coincide. Let

$$T_n = \inf\{t \geq 0 \, : \, \langle M \rangle_t, \langle N \rangle_t |A|_t \text{ or } |B|_t \geq n\}.$$

If H is locally bounded and predictable and $H_t = 0$ for $t \geq T_n$ we can find simple H^m such that

$$\|H^m - H\|_M, \|H^m - H\|_N \to 0 \text{ and } \int |H_s^m - H_s| dV_s \to 0,$$

where V_t is the sum of the variation of A and B on $[0, t]$. We can now take the equality

$$(H^m \cdot (M + A))_t = (H^m \cdot (N + B))_t$$

and let $m \to \infty$. For the left hand side we have

$$\lim_{m \to \infty} \sup_t \int \Big((H^m \cdot (M + A))_t - (H \cdot (M + A))_t \Big)^2 dP = 0.$$

In particular, we can find a subsequence, such that the left hand side converges almost surely [P_t] to $(H \cdot X)_t$ and for this subsequence we find a further subsequence such that the right hand side converges almost surely[Q_t] to $(H \cdot X)_t$, now calculated with respect to Q. This also implies the statement, because almost surely[P_t] and almost surely[Q_t] are the same. ■

Example: Look at a Brownian motion with constant drift $X : t \mapsto X_t = B_t + bt$. We should first find out, whether the distribution Q of X is locally equivalent to Wiener's measure $P = \mathbf{P}_0$. This is statistically relevant: if the answer is yes, it is impossible distinguish Brownian motion X with drift and B without drift with 100% certainty in any finite time interval.

Let $0 = t_0 < t_1 < t_2 < \cdots < t_n = t$ and calculate (with $y_0 := 0$)

$$\begin{aligned}
&\mathbf{P}_0\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\} \\
&= \int_{A_1} dy_1 \cdots \int_{A_n} dy_n \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}^d} \exp\Big(- (y_i - bt_i - y_{i-1} + bt_{i-1})^2/2(t_i - t_{i-1})\Big) \\
&= \int_{A_1} dy_1 \cdots \int_{A_n} dy_n \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}^d} \\
&\quad \times \exp\Big(- (y_i - y_{i-1})^2/2(t_i - t_{i-1})\Big) \exp\Big(b(y_i - y_{i-1})\Big) \exp\Big(- b^2(t_i - t_{i-1})/2\Big) \\
&= \mathbf{E}_0\Big\{ \prod_{i=1}^n 1_{A_i}(B(t_i)) \exp\Big(bB_t - \frac{b^2}{2}t\Big)\Big\}
\end{aligned}$$

Using the monotone class theorem we can infer, for every bounded $\mathcal{F}_t$-measurable function g,

$$\mathbb{E}_0\Big\{g(X)\Big\} = \mathbf{E}_0\Big\{g(B) \exp\Big(bB_t - \frac{b^2}{2}t\Big)\Big\}.$$

We can infer that P and Q are locally equivalent with

$$\alpha_t(f) = \exp\Big(bf(t) - \frac{b^2}{2}t\Big).$$

Hence the process $\{\alpha_t\}$ under Wiener's measure $\mathbf{P}_0$ is a geometric Brownian motion (and by Theorem 1.4 a martingale!). Girsanov tells us now, if X is a Brownian motion with drift, the quantity A_t we have to subtract to get a local martingale[Q] is

$$A_t = \int_0^t \exp\Big(- bX(s) + \frac{b^2}{2}s\Big) d\Big\langle \exp\Big(bX(t) - \frac{b^2}{2}t\Big), X\Big\rangle_s.$$

This should be equal to bt, let's check it. By Ito's formula, we have

$$\exp\Big(bX(t) - \frac{b^2}{2}t\Big) - \exp\Big(-\frac{b^2}{2}t\Big) = \int_0^t b\exp\Big(bX(s) - \frac{b^2}{2}s\Big)\, dX_s + \frac{1}{2}\int_0^t b^2 \exp\Big(bX(s) - \frac{b^2}{2}s\Big)\, ds.$$

Hence, we have

$$\Big\langle \exp\Big(bX(t) - \frac{b^2}{2}t\Big), X\Big\rangle_t = \Big\langle \int_0^t b\exp\Big(bX(s) - \frac{b^2}{2}s\Big)\, dB_s, B\Big\rangle_t = b\int_0^t \exp\Big(bX(s) - \frac{b^2}{2}s\Big)\, ds.$$

Altogether,

$$A_t = b\int_0^t \exp\Big(- bX(s) + \frac{b^2}{2}s\Big) \exp\Big(bX(s) - \frac{b^2}{2}s\Big)\, ds = bt.$$

This example is just the simplest in a long row. One can make the drift dependent on space and time, etc. A question worth asking is for which other function $f(t)$ the shifted Brownian motion $B(t) + f(t)$ has a distribution locally equivalent to Brownian motion without drift. This applies if f is in the *Cameron-Martin space*

$$\Big\{f : [0,\infty) \to \mathbf{R} \,:\, \exists c : [0,\infty) \to \mathbf{R} \text{ measurable }, f(t) = \int_0^t c(s)\, ds \text{ and } \int_0^t c^2(s)\, ds < \infty.\Big\}.$$

Define $Y(t) = \int_0^t c(s)\, dB_s$. This is well-defined, because $c \in \Pi_3(B)$ by the square-integrability assumption. Then

$$\langle Y \rangle_t = \int_0^t c^2(s)\, ds.$$

In this case we have,

$$\alpha_t := \exp\left(Y_t - \frac{1}{2}\langle Y \rangle_t\right)$$

To check this one can calculate, for a $\mathcal{F}_t$-measurable function g,

$$\mathbf{E}_0\Big\{g(B) \exp\left(Y_t - \frac{1}{2}\langle Y \rangle_t\right)\Big\} = \mathbf{E}_0\Big\{g(B + f)\Big\}.$$

At least heuristically this can be seen using the same approach as in the case of constant drift. In the theory of stochastic differential equations Girsanov's formula is used to construct solutions of SDE's using the change of measure technique.

Chapter 5

The Meyer-Tanaka formula and applications

In this chapter we generalise Itô's formula by relaxing the differentiability condition on f. The resulting formula, the Meyer-Tanaka formula, will be applied to two problems,

- constructing a family L^a of processes, which measure in a certain sense how much time a one-dimensional Brownian motion $\{B(t) : t \in [0,1]\}$ spends at a given point a. L^a is the so-called *local time at* a,

- finding the law of the time spent by a Brownian motion $\{B(t) : t \in [0,1]\}$ above the x-axis. This will give Paul Lévy's famous arcsine law.

5.1 The Meyer-Tanaka formula

The Meyer-Tanaka formula extends the Itô formula to a class of functions $f : \mathbb{R} \to \mathbb{R}$, which are not necessarily twice continuously differentiable, namely the class of *convex functions*. The function f is called *convex* if

$$f(ax + (1-a)y) \le af(x) + (1-a)f(y) \text{ for all } a \in [0,1] \text{ and } x, y \in \mathbb{R}.$$

f need not be differentiable, but we can define objects which replace the first and second derivatives.

Proposition 5.1 *Every convex function has a left derivative*

$$f'(x) = \lim_{h \downarrow 0} \frac{f(x) - f(x-h)}{h} \text{ at all } x \in \mathbb{R}.$$

f' is left continuous and increasing and hence there is a locally finite measure μ such that $f'(x) - f'(y) = \mu[y, x)$. In other words, f' is the distribution function of the measure μ.

Proof: Using convexity we infer that, for all $h < k$,

$$\begin{aligned} \frac{f(y)-f(y-h)}{h} - \frac{f(y)-f(y-k)}{k} &= \frac{1}{h}\Big((1-\frac{h}{k})f(y) + \frac{h}{k}f(y-k) - f(y-h)\}\Big) \\ &\geq \frac{1}{h}\Big(f(y-h) - f(y-h)\Big) = 0. \end{aligned}$$

Hence the differential quotient is increasing as $h \downarrow 0$ and must have a limit $f'(y) \in \mathbb{R} \cup \{\infty\}$, such that f' is a lower semicontinuous function. To show that f' is increasing let $x_1 < x_2$ and $x = \frac{1}{2}(x_1 + x_2)$ the midpoint. Then

$$\frac{1}{2}f(x_2) + \frac{1}{2}f(x_1) \geq f(x) = \frac{1}{2}f(x) + \frac{1}{2}f(x),$$

and it follows that $f(x_2) - f(x) \geq f(x) - f(x_1)$. Iterating the argument one can see that, whenever $x > y$ and $h_n = (y-x)/2^n$,

$$f(x) - f(x-h_n) \geq f(y) - f(y-h_n).$$

Dividing by h_n and taking limits gives the statement. Now it is clear that $f'(y) < \infty$ for all y. Note that every increasing, lower semicontinuous function f' is left continuous and from this it is easy to see that a locally finite measure μ can be defined uniquely by $\mu[a,b) = f'(b) - f'(a)$.
■

If g is the distribution function of μ, then we say that $g' = \mu$ in the distributional sense, so that in the last proposition $f'' = \mu$ in the distributional sense. The main result of this section is a formula, which shows that $\{f(X_t) : t \geq 0\}$ is again a semimartingale.

Theorem 5.2 (Meyer-Tanaka Formula) *Let X be a continuous semimartingale. Define a family $L^a = \{L^a_t : t \geq 0\}$, for $a \in \mathbb{R}$, of increasing processes by*

$$L^a_t = |X_t - a| - |X_0 - a| - \int_0^t \text{sign}(X_s - a)\, dX_s, \tag{5.1}$$

where sign $x = 1_{\{x>0\}} - 1_{\{x\leq 0\}}$. *Then, for every convex function f with left derivative f' and $f'' = \mu$ in the distributional sense, we have*

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)\, dX_s + \frac{1}{2}\int L^a_t\, d\mu(a).$$

Remarks: Clearly, if f is the difference of two convex functions, we still have that $\{f(X_t) : t \geq 0\}$ is a semimartingale. This result is optimal in the sense that, if B is standard Brownian motion and $\{f(B_t) : t \geq 0\}$ is a semimartingale, then f must be the difference of two convex functions, by a theorem of Cinlar, Jacod, Protter and Sharpe (1980).

Now look at a locally finite measure μ. Define a convex function

$$f(x) = \begin{cases} \int_0^x \mu[0,y)\, dy & \text{if } x > 0, \\ \int_x^0 \mu[y,0)\, dy & \text{if } x \leq 0. \end{cases}$$

Then $f'' = \mu$ in the distributional sense. In the special case that $\mu = \delta_a$, say for some $a \geq 0$, we have $f'(x) = 1_{(a,\infty)}(x)$ and

$$f(x) = (x - a)^+ \text{ for all } x \in \mathbb{R}.$$

If we formally compared the Meyer-Tanaka formula for these f and Itô's formula for f, we would get

$$L_t^a = \int_0^t \delta_a(X_s)\, d\langle X\rangle_s.$$

This is of course not rigourous, because f'' is taken in the distributional sense only, but it indicates that L_t^a measures the time that the process X spends in a with respect to the clock given by the process $\langle X\rangle$. This justifies that the process L^a is called *the local time of* X *at* a. We will make this interpretation precise in the next proposition and, in particular, in the next section.

For the proof of the Meyer-Tanaka formula we proceed in two steps. We first show a less explicit version of the result.

Lemma 5.3 (Step 1) *Let X be a continuous semimartingale and f a convex function. Then there exists a continuous, adapted increasing process $K[f] := \{K_t : t \geq 0\}$ such that*

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)\, dX_s + K_t.$$

We start with the **Proof of Step 1**. Let $X = M + A$ be the decomposition of the semimartingale. By stopping we can assume that $|X_t|$, $|A_t|$ and $\langle M\rangle_t \leq N$ for all t. Let g be a nonnegative C^∞ function with compact support in $(-\infty, 0]$ and $\int g(s)\, ds = 1$. The idea is to use g to make f smooth. Let

$$f_n(x) = \int f\Big(x - \frac{y}{n}\Big) g(y)\, dy = n \int f(x - z) g(nz)\, dz.$$

First, f_n is the convolution of f with the smooth function $ng(nz)$ and therefore it is smooth. It is easy to check that it inherits convexity from f, so that f_n is both convex and C^∞. Hence we can apply Itô's formula to get

$$f_n(X_t) - f_n(X_0) = \int_0^t f_n'(X_s)\, dX_s + \frac{1}{2}\int_0^t f_n''(X_s)\, d\langle X\rangle_s. \tag{5.2}$$

The plan is to let $n \to \infty$ to get the formula. The sequence of probability measures $ng(nz)\, dz$ converges weakly to δ_0, which implies that $f_n(x) \to f(x)$ for all x. As a convex function, f is Lipschitz continuous on compact intervals and from the definition of f_n we infer that on each compact interval all functions f_n are Lipschitz with a universal Lipschitz constant. Hence, given $\varepsilon > 0$ and a compact interval $I = [a, b]$, we can find a Lipschitz constant C and an N such that $|f_n(x_k) - f(x_k)| < \varepsilon/2$ for all $n \geq N$ and $x_k = a + k\varepsilon/4C \leq b$. Then, for all $x \in I$ there is $x_k \leq x \leq x_{k+1}$ and

$$\begin{aligned} |f_n(x) - f(x)| &\leq |f_n(x_k) - f_n(x)| + |f_n(x_k) - f(x_k)| + |f(x_k) - f(x)| \\ &\leq 2C(\varepsilon/4C) + \varepsilon/2 \leq \varepsilon. \end{aligned}$$

In other words, $f_n(x) \to f(x)$ uniformly on compact intervals. As $|X_t| \leq N$ we have

$$f_n(X_t) \to f(X_t) \text{ uniformly in } t.$$

This ensures uniform convergence of the left hand side. To deal with the right hand side, we differentiate f_n and get

$$f_n'(x) = \int f'\Big(x - \frac{y}{n}\Big) g(y)\, dy \uparrow f'(x) \text{ as } n \to \infty.$$

Now let

$$I_t^n := \int_0^t f_n'(X_s)\, dM_s \text{ and } I_t := \int_0^t f'(X_s)\, dM_s.$$

We use Doob's L^2 maximal inequality and the isometry property Theorem 2.23 to infer

$$\begin{aligned} \mathbf{E}\Big\{ \sup_t (I_t^n - I_t)^2 \Big\} &\le 4 \sup_t \mathbf{E}\{(I_t^n - I_t)^2\} \\ &= 4\mathbf{E} \int_0^\infty (f_n'(X_s) - f'(X_s))^2\, d\langle M\rangle_s \longrightarrow 0, \end{aligned}$$

by the bounded convergence theorem. By passing to a subsequence one can get

$$\sup_t (I_t^n - I_t)^2 \longrightarrow 0 \text{ almost surely.}$$

Now let

$$J_t^n := \int_0^t f_n'(X_s)\, dA_s \text{ and } J_t := \int_0^t f'(X_s)\, dA_s.$$

From bounded convergence it follows that, for almost every ω,

$$\sup_t |J_t^n - J_t| \le \int_0^\infty |f_n'(X_s) - f'(X_s)|\, dA_s \to 0.$$

By now we have shown that

$$K_t^n := \frac{1}{2} \int_0^t f_n''(X_s)\, d\langle X\rangle_s = f_n(X_t) - f_n(X_0) - \int_0^t f_n'(X_s)\, dX_s$$

has a subsequence, which converges uniformly in t. We denote the limit K_t and recall that the uniformity of the convergence implies that the limit process is continuous, adapted and increasing. ∎

Note that at this point we already have that $\{f(X_t) : t \ge 0\}$ is a semimartingale. If we let $L^a := 2K[(x-a)^+]$, then by Step 1,

$$\frac{1}{2} L_t^a = (X_t - a)^+ - (X_0 - a)^+ - \int_0^t 1_{\{X_s > a\}}\, dX_s,$$

so that our definition is good for the special choice $f(x) = (x-a)^+$. It is not hard to see, assuming the result of Step 1, that this definition of L^a agrees with the one given in the Meyer-Tanaka Theorem.

Lemma 5.4 $K[|\cdot - a|] = 2K[(\cdot - a)^+] = 2K[(\cdot - a)^-]$.

Proof: Let $f_1(x) = (x-a)^+$, $f_2(x) = (x-a)^-$ and $f(x) = |x-a|$, so that we have to show that the processes $K[f_1] = K[f_2] = \frac{1}{2}K[f]$. Observe that $f_1 + f_2 = f$ and hence, by uniqueness $K[f] = K[f_1] + K[f_2]$. As $f_1 - f_2 = x - a$ we have $K[f_1] - K[f_2] = 0$ and hence we have $K[f_1] = K[f_2] = \frac{1}{2}K[f]$, as required. ■

Now observe that $f_2' = -1_{(-\infty,a]}$ and infer by adding Step 1 for f_1 and f_2 that the two definitions of L^a agree. We are going to use only Step 1 for the proof of the following proposition, which makes part of our intuition about local times L^a rigourous.

Proposition 5.5 *For every a, almost surely, the process $L^a = \{L^a_t : t \geq 0\}$ increases only on the set $\{t : X_t = a\}$. More formally, let ℓ^a be the measure with distribution function L^a, then ℓ^a is supported by the level set $\{X_t = a\}$.*

Proof: Let $S < T$ be stopping times such that $[S,T] \subset \{X_t < a\}$. Applying Step 1 gives

$$(X_T - a)^+ - (X_S - a)^+ = \int_S^T 1_{\{X_s > a\}}\, dX_s + \frac{1}{2}(L^a_T - L^a_S).$$

The left hand side and the integral are zero, so that $L^a_T = L^a_S$. This holds, almost surely, for all stopping times

$$S = \inf\{t > q \,:\, X_t \leq a - \frac{1}{m}\} \text{ and } T = \inf\{t > S \,:\, X_t \geq a - \frac{1}{n}\},$$

for all q rational and $m < n$ positive integer. It follows that

$$\ell^a\{t \,:\, X_t < a\} = 0$$

and one can argue similarly with f_2 replacing f_1 to infer $\ell^a\{t : X_t > a\} = 0$. ■

In the second step we prove a lemma, which gives the complete connection between $K[f]$ for different f.

Lemma 5.6 (Step 2) *If $f'' = \mu$, then $K[f] = \{\frac{1}{2}\int L^a_t \, d\mu(a) \,:\, t \geq 0\}$.*

It should now be clear that the Meyer-Tanaka formula follows from the two steps and we proceed to the **proof of Step 2**. For this purpose we need a step of independent interest, a Fubini's Theorem that allows to interchange stochastic and ordinary integration.

Theorem 5.7 (Fubini's Theorem) *Suppose μ is a finite measure and X a continuous semi-martingale and H bounded and measurable with respect to $\mathcal{B} \otimes \Pi$, the product of the Borel σ-field on $\mathbf{R}$ and the predictable σ-field on $[0,\infty) \times \Omega$. Consider $H^a = H(a,\cdot)$ as an integrand parametrised by $a \in \mathbf{R}$. There exists a $\mathcal{B} \otimes \Pi$ measurable random variable Z such that for μ-almost every a the process $Z^a = Z(a,\cdot)$ is a continuous version of the process $\int_0^t H^a_s \, dX_s$. Let*

$$Y_t = \int Z^a_t \, \mu(da) \text{ and } J_t = \int H^a_t \, \mu(da).$$

Then Y is a version of $J \cdot X$, or, less formally,

$$\int \int_0^t H^a_s \, dX_s \, \mu(da) = \int_0^t \int H^a_s \, \mu(da) \, dX_s.$$

We first use this result to do Step 2 of the Meyer-Tanaka formula, then we sketch a proof of Fubini's Theorem.

Proof of Step 2: By stopping we can assume $|X_t| \leq N$ and $\langle X \rangle_t \leq N$ for all t. Let

$$g(x) = \frac{1}{2} \int_{-N}^{N} |x - a| \, d\mu(a).$$

g is a convex function with

$$g'(x) = \frac{1}{2} \int_{-N}^{N} \text{sign}(x - a) \, d\mu(a)$$

and $g''(x) = \mu$ for all $|x| \leq N$. Hence $f(x) - g(x) = a + bx$ for $|x| \leq N$. Note that the Meyer-Tanaka formula holds trivially for the linear function $a + bx$ and thus if suffices to prove it for g. Starting with the definition of local time

$$L_t^a = |X_t - a| - |X_0 - a| - \int_0^t \text{sign}(X_s - a) \, dX_s$$

and integrating with respect to $\frac{1}{2} \int_{-N}^{N} d\mu(a)$ gives

$$g(X_t) - g(X_0) = \frac{1}{2} \int_{-N}^{N} \int_0^t \text{sign}(X_s - a) \, dX_s \, d\mu(a) + \frac{1}{2} \int_{-N}^{N} L_t^a \, d\mu(a).$$

By Fubini's Theorem 5.7 we can interchange the order of stochastic and Lebesgue integration in the first term on the right, and get

$$g(X_t) - g(X_0) = \int_0^t g'(X_s) \, dX_s + \frac{1}{2} \int_{-N}^{N} L_t^a \, d\mu(a).$$

Now recall that we have stopped X upon leaving $[-N, N]$, so that we can infer from Lemma 5.5 that $L_t^a = 0$ for all $|a| \geq N$. Hence we can extend the last integral to the whole real line without changing its value. ■

Proof of Fubini's Theorem: We leave out one technical detail from the proof (roughly half a page of proof, see Durrett p.86), the existence of a process Z, which is measurable on $\mathcal{B} \otimes \Pi$, such that, for μ-almost every a, the process $Z^a = Z(a, \cdot)$ is a continuous version of $\int_0^t H_s^a \, dX_s$. Note that only measurability is the issue here, we already know that for every single a a continuous version of $\int_0^t H_s^a \, dX_s$ exists.

For the proof let $X = M + A$ be the semimartingale decomposition and assume, by stopping, that $|A_t| \leq N$, $|M_t| \leq N$ and $\langle X \rangle_t = \langle M \rangle_t \leq N$ for all t. The case that $X = A$ is the ordinary Fubini's Theorem so that we can assume that $X = M$. The idea of proof is to apply the monotone class theorem to the collection $\mathcal{H}$ of bounded $\mathcal{B} \otimes \Pi$-measurable functions, for which the statement holds true. It is clear that $\mathcal{H}$ is a vector space and we first check that the indicator functions of sets from the collection

$$\mathcal{A} = \{A \times B \, : \, A \in \mathcal{B}, B \in \Pi\}$$

which is a $\cap$-stable generator of $\mathcal{B}\otimes\Pi$, are all in the class $\mathcal{H}$. Indeed, suppose $H(a,t,\omega) = 1_A(a)1_B(t,\omega)$. Then $Z_t^a = 1_A(a)\int_0^t 1_B(s,\omega)\,dX_s$ and

$$\begin{aligned}\int Z_t^a\,d\mu(a) &= \int_0^t 1_B(s,\omega)\,dX_s \int 1_A(a)\,d\mu(a)\\ &= \int_0^t \Big(\int 1_A(a)\,d\mu(a)1_B(s,\omega)\Big)\,dX_s\\ &= (J\cdot X)_t.\end{aligned}$$

It remains to check that for $0\le H^n \uparrow H$ with $H^n\in\mathcal{H}$ and H bounded, we can infer that $H\in\mathcal{H}$. For this purpose let $\|\mu\|$ be the total mass of μ and apply Jensen to $\mu/\|\mu\|$, so that

$$\frac{1}{\|\mu\|}\mathbf{E}\Big\{\Big(\int \sup_t |Z_t^{n,a}-Z_t^a|\,d\mu(a)\Big)^2\Big\} \le \mathbf{E}\Big\{\int \sup_t |Z_t^{n,a}-Z_t^a|^2\,d\mu(a)\Big\}.$$

Using the ordinary Fubini's Theorem, the L^2-maximal inequality, and the isometry property we get for the right hand side

$$\begin{aligned} &= \int \mathbf{E}\{\sup_t |Z_t^{n,a}-Z_t^a|^2\}\,d\mu(a)\\ &\le 4\int \sup_t \mathbf{E}|Z_t^{n,a}-Z_t^a|^2\,d\mu(a)\\ &= 4\int \mathbf{E}\Big\{\int_0^\infty (H_s^{n,a}-H_s^a)^2\,d\langle X\rangle_s\Big\}\,d\mu(a)\longrightarrow 0,\end{aligned}$$

by iterated use of bounded convergence. We can infer from this that

$$\mathbf{E}\Big\{\sup_t\Big|\int Z_t^{n,a}\,d\mu(a)-\int Z_t^a\,d\mu(a)\Big|\Big\}^2 \le \mathbf{E}\Big\{\Big(\int \sup_t |Z_t^{n,a}-Z_t^a|^2\,d\mu(a)\Big)\Big\}\longrightarrow 0.$$

Taking $J_t^n = \int H_t^{n,a}\,d\mu(a)$, we have that, almost surely, a subsequence of $(J^n\cdot X)_t = \int Z_t^{n,a}\,d\mu(a)$ converges uniformly to $\int Z_t^a\,\mu(da)$. It remains to check that $J^n\cdot X$ converges to $J\cdot X$ in $\mathcal{M}^2$, because this implies $\int Z_t^a\,d\mu(a) = (J\cdot X)_t$. By the isometry property this is equivalent to showing that $\|J^n - J\|_X \to 0$. This follows from

$$\frac{1}{\|\mu\|}\mathbf{E}\Big\{\int_0^\infty\Big(\int H_s^{n,a}\,d\mu(a)-\int H_s^a\,d\mu(a)\Big)^2\,d\langle X\rangle_s\Big\} \le \mathbf{E}\Big\{\int_0^\infty\int (H_s^{n,a}-H_s^a)^2\,d\mu(a)\,d\langle X\rangle_s\Big\}\to 0\,,$$

where we have used the bounded convergence theorem three times. This proves Fubini's Theorem for H and, by application of the monotone class theorem, the statement. ∎

5.2 Local times of Brownian motion

We now concentrate on a (one-dimensional) Brownian motion B. In the previous section we have seen that the processes $L^a = \{L_t^a : t\ge 0\}$ increases only on the set $\{B_t = a\}$. The following theorem makes clear in which sense L^a is the time spent by Brownian motion in a.

Theorem 5.8 *Suppose B is a one-dimensional Brownian motion and let*

$$\mu_t(A) = \int_0^t 1_A(B_s)\, ds \text{ for } A \subset \mathbb{R} \text{ Borel,}$$

be the occupation measure of Brownian motion, i.e. the (random) measure assigning to each set A the time spent by Brownian motion up to time t in this set. Then, almost surely, the measure μ_t has a density with respect to Lebesgue measure given by $a \mapsto L_t^a$. Equivalently,

$$\int_{-\infty}^{\infty} L_t^a\, g(a)\, da = \int_0^t g(B_s)\, ds = \int g\, d\mu_t,$$

for all bounded measurable functions g.

Proof: It suffices, by the monotone class theorem, to prove this for continuous $g \geq 0$. We then find f twice continuously differentiable such that $f'' = g$ and this f must be convex. So we can equate Itô's formula and the Meyer-Tanaka formula for $f(B_t) - f(B_0)$ and get that,

$$\int L_t^a\, g(a)\, da = \int_0^t f''(B_s)\, ds = \int_0^t g(B_s)\, ds = \int g\, d\mu_t.$$

■

Remark: It can be shown that the mapping $(a, t) \mapsto L_t^a$ can be chosen continuously, so that the formula in the theorem holds.

By a measure-theoretic theorem of Lebesgue one can get the density of any absolutely continuous measure by considering the limit of a shrinking family of balls. In the situation of Theorem 5.8 this gives the following representation of local times.

Corollary 5.9 *For every t, almost surely, for Lebesgue-almost every a,*

$$L_t^a = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^t 1_{(a-\varepsilon, a+\varepsilon)}(B_s)\, ds.$$

Remark: Using continuity of $(a, t) \mapsto L_t^a$ one can get this representation almost surely for every a and t.

Local times can be considered as a processes in the space variable a or in the time variable t. Both points of view give interesting results. We mention two of them (without proof).

Theorem 5.10 (Lévy's Theorem) *The processes $t \mapsto L_t^0$ and $t \mapsto \max_{0 \leq s \leq t} B_s$ have the same distribution.*

Note on the proof: One has to use that a continuous local martingale M with $\langle M \rangle_t = t$ is a Brownian motion (this is another famous result of Lévy). It follows that the process

$$W_t = \int_0^t \text{sign}(B_s)\, dB_s$$

is a standard Brownian motion and, by Meyer-Tanaka, we have

$$|B_t| = W_t + L_t^0.$$

One then uses a lemma of Skorokhod stating that every continuous function $f : [0,\infty) \to \mathbf{R}$ with $f(0) \geq 0$ can be written uniquely as $f(t) = z(t) - a(t)$ where z is positive on $(0,\infty)$ and a is increasing, continuous, vanishing at 0 such that the measure da_s is supported by $\{s : z(s) = 0\}$. Then always

$$a(t) = \sup_{s\leq t}\{-f(s) \vee 0\}.$$

Applying this deterministic result to $t \mapsto W_t$ gives

$$L_t^0 = \sup_{s\leq t}\{-W_s\} \text{ for } t > 0.$$

This is the statement, considering that $-W_t$ is a standard Brownian motion. See Revuz/Yor (p. 229) for details.

We finish the section by mentioning the following exciting result of Ray and Knight (1963) about local time as a process in the space variable.

Theorem 5.11 (Ray-Knight Theorem) *Suppose $b > 0$ and $T = \inf\{t > 0 : B_t = b\}$. Then the distribution of the process $\{L_T^{b-x} : x \in [0,b]\}$ agrees with the distribution of $\{\frac{1}{2}\|X_t\|^2 : t \in [0,b]\}$, where X is a two-dimensional Brownian motion.*

The Meyer-Tanaka formula plays an important role in the proof of this result, which we cannot give here for time reasons, see e.g. Revuz-Yor p.434.

5.3 Paul Lévy's arcsine law

Another application of the Meyer-Tanaka formula is Kac's proof of Paul Lévy's famous arcsine law. Let $H(t)$ be the proportion of time in $[0,t]$ spent by Brownian motion above the x-axis. What is the distribution of H_t/t?

Theorem 5.12 (Arcsine Law) *For all $x \in [0,1]$,*

$$\mathbf{P}_0\{H_t \leq xt\} = \frac{1}{\pi}\int_0^x \frac{dr}{\sqrt{r(1-r)}} = \frac{2}{\pi}\arcsin(\sqrt{x}).$$

Remarks:

- Note that the distribution of H_t/t does not depend on t and is a distribution with maximum of the density function near 0 and 1. Heuristically this means, that the proportion of time a Brownian gambler is ahead does not converge to 1/2 but fluctuates such that most of the time it is rather nearer to 0 or 1.

- There is a brother to this theorem, the arcsine law for the last sign change of a Brownian motion, which is easier to prove, see Step 4 in the proof of Theorem 6.9 in Probability Theory.

- The second equality can be checked by differentiating arcsin.
- Donsker's invariance principle can be used to get a result like Theorem 5.12 for random walks. The argument is like the first three steps in the proof of Theorem 6.9 in Probability Theory. This is a recommended exercise.

For the proof we need a probabilistic representation of the solutions of a special differential equations, but now the solutions are just in a distributional sense.

Lemma 5.13 *Let $c(x) = -\alpha - \beta 1_{[0,\infty)}(x)$ where $\alpha, \beta > 0$. Suppose v is a bounded C^1-function, which satisfies*

$$\frac{1}{2}v''(x) + c(x)v(x) = -1$$

in the distributional sense, then for all x,

$$v(x) = \int_0^\infty e^{-\alpha t} \mathbf{E}_x \Big\{ e^{-\beta H_t} \Big\} \, dt.$$

Remark: The equation in the distributional sense means that the function v' is the difference of the distribution functions of the measures whose densities are the positive and the negative part of $-2(1 + c(x)v(x))$. Note that we cannot expect to find a C^2 solution because c is not continuous at 0.

Proof: Note that v is the difference of two convex functions. By the Meyer-Tanaka formula and Theorem 5.8

$$v(B_t) - v(B_0) = \int_0^t v'(B_s)\, dB_s - \int (1 + c(a)v(a)) L_s^a \, da = \int_0^t v'(B_s)\, dB_s - \int_0^t (1 + c(B_s)v(B_s))\, ds.$$

Letting $c_t = \int_0^t c(B_s)\, ds$ and using the integration by parts formula with $X_t = v(B_t)$ and $Y_t = \exp(c_t)$, which is locally of bounded variation, we have

$$\begin{aligned} v(B_t)\exp(c_t) - v(B_0) &= \int_0^t \exp(c_s) v'(B_s)\, dB_s - \int_0^t \exp(c_s)(1 + c(B_s)v(B_s))\, ds \\ &\quad + \int_0^t v(B_s) \exp(c_s)\, dc_s. \end{aligned}$$

By definition of c_t we have

$$\int_0^t \exp(c_s) c(B_s) v(B_s)\, ds = \int_0^t v(B_s) \exp(c_s) dc_s,$$

and hence $M_t = v(B_t)\exp(c_t) + \int_0^t \exp(c_s)\, ds$ defines a local martingale. Since v is bounded and $c_t \le -\alpha < 0$, M is bounded. As $t \to \infty$, $\exp(c_t) \le e^{-\alpha t} \to 0$, and by martingale convergence

$$v(x) = \mathbf{E}_x\{M_0\} = \mathbf{E}_x\{M_\infty\} = \mathbf{E}_x \int_0^\infty \exp(c_s)\, ds = \mathbf{E}_x \int_0^\infty \exp\Big(\int_0^t c(B_s)\, ds \Big)\, dt,$$

using also bounded convergence. Now recall the definition of c and use Fubini's Theorem to get that

$$\mathbf{E}_x \int_0^\infty \exp\Big(\int_0^t c(B_s)\, ds \Big)\, dt = \int_0^\infty e^{-\alpha t} \mathbf{E}_x \Big\{ e^{-\beta \int_0^t 1_{[0,\infty)}(B_s)\, ds} \Big\},$$

which is the statement. ■

Our strategy now is to guess a solution v of the problem in Lemma 5.13 to get a formula for $\mathbf{E}_0\{\exp(-\beta H_t)\}$. For this purpose one can find a C^1 function v, which is C^2 on $\mathbf{R}\setminus\{0\}$ such that

$$\begin{array}{ll} (\alpha+\beta)v = \frac{1}{2}v'' + 1 & \text{if } x > 0 \\ \alpha v = \frac{1}{2}v'' + 1 & \text{if } x < 0. \end{array}.$$

One easily (with a bit of experience) finds a candidate

$$v(x) := \begin{cases} A\exp(-x\sqrt{2(\alpha+\beta)}) + \frac{1}{\alpha+\beta} & \text{if } x > 0, \\ B\exp(x\sqrt{2\alpha}) + \frac{1}{\alpha} & \text{if } x \le 0, \end{cases}$$

with

$$A = \frac{\sqrt{\alpha+\beta}-\sqrt{\alpha}}{(\alpha+\beta)\sqrt{\alpha}} \text{ and } B = \frac{\sqrt{\alpha}-\sqrt{\alpha+\beta}}{\sqrt{\alpha+\beta}\alpha}.$$

For the proof, an easy calculation confirms that this is a solution of the problem on $\mathbf{R}\setminus\{0\}$. Because v' is continuous in 0 it must be the distribution function of $-2(1+cv)$. We infer that

$$\frac{1}{\sqrt{\alpha(\alpha+\beta)}} = v(0) = \int_0^\infty e^{-\alpha t}\mathbf{E}_0\Big\{e^{-\beta H_t}\Big\}\,dt.$$

The right hand side is the Laplace transform of $t \mapsto \mathbf{E}_0\{e^{-\beta H_t}\}$. It would be good to write the left hand side as a Laplace transform, too, because then we can apply the fact that the Laplace transform uniquely determine the function. For this purpose observe that

$$\int_0^\infty \frac{e^{-\gamma t}}{\sqrt{t}}\,dt = \sqrt{\frac{2}{\gamma}}\int_0^\infty e^{-x^2/2}\,dx = \sqrt{\frac{\pi}{\gamma}}.$$

Hence,

$$\frac{1}{\sqrt{\alpha(\alpha+\beta)}} = \frac{1}{\pi}\int_0^\infty \frac{e^{-(\alpha+\beta)s}}{\sqrt{s}}\int_s^\infty \frac{e^{-\alpha(t-s)}}{t-s}\,dt\,ds = \frac{1}{\pi}\int_0^\infty e^{-\alpha t}\int_0^\infty \frac{e^{-\beta s}}{\sqrt{s(t-s)}}\,ds\,dt.$$

From the uniqueness of Laplace transforms we get

$$\mathbf{E}_0\{\exp(-\beta H_t)\} = \frac{1}{\pi}\int_0^t \frac{e^{-\beta s}}{\sqrt{s(t-s)}}\,ds$$

and using uniqueness of Laplace transforms of probability distributions we get the desired result. ■

Bibliography

The lecture was based on the following books:

- Richard Durrett, “Stochastic Calculus” CRC Press, Boca Raton.
- Heinrich v. Weizsäcker and Gerhard Winkler, “Stochastic Integrals” Vieweg: Series Advanced Lectures in Mathematics.
- Ioannis Karatzas and Steven E. Shreve “Brownian motion and Stochastic Calculus” Springer: Series Graduate Texts.
- Richard Bass, “Probabilistic Techniques in Analysis” Springer: Series Probability and its Applications.
- Daniel Revuz and Marc Yor, “Brownian motion and continuous martingales” Springer: Series Grundlehren.

Recommended further reading are the second half of Durrett’s book concerning stochastic differential equations and, in particular, the many aspects covered in Bass’ book which for time reasons couldn’t be touched in the lecture.