or in integral form

$$f\big(x(t)\big) - f\big(x(0)\big) = \int_0^t f'\big(x(s)\big)\,\mathrm{d}x(s). \tag{1.2}$$

In this last form, the chain rule can be extended to the case where f is in C^1 and x is *continuous* and of *finite variation*.

Unfortunately, this classical result does not help us a lot. For one thing, X might have only RCLL instead of continuous trajectories. This is still solvable if X has trajectories of finite variation. But even if X is continuous, we cannot hope that its trajectories are of finite variation, as the example of X being a Brownian motion clearly demonstrates. So we need a different result, namely a chain rule for functions having a nonzero quadratic variation.

Let us now connect the above idea to semimartingales. Recall that a *semimartingale* is a stochastic process of the form $X = X_0 + M + A$, where M is a local martingale null at 0 and A is an adapted process null at 0 with RCLL trajectories of finite variation. For any such A and any fixed, i.e. nonrandom, sequence $(\Pi_n)_{n\in\mathbb{N}}$ of partitions of $[0,\infty)$ with $\lim_{n\to\infty}|\Pi_n| = 0$, the *quadratic variation* of A along $(\Pi_n)_{n\in\mathbb{N}}$ is given by the sum of the squared jumps of A, i.e.

$$[A]_t = \lim_{n\to\infty} \sum_{t_i\in\Pi_n} (A_{t_{i+1}\wedge t} - A_{t_i\wedge t})^2 = \sum_{0<s\le t} (\Delta A_s)^2 = \sum_{0<s\le t} (A_s - A_{s-})^2 \qquad \text{for } t \ge 0.$$

By polarisation, we then obtain for any semimartingale Y that

$$[A, Y]_t = \sum_{0<s\le t} \Delta A_s \Delta Y_s \qquad \text{for } t \ge 0.$$

So the quadratic variation of a general semimartingale $X = X_0 + M + A$ has the form

$$[X] = [M + A] = [M] + [A] + 2[M, A] = [M] + \sum_{0<s\le\cdot} (\Delta A_s)^2 + 2 \sum_{0<s\le\cdot} \Delta M_s \Delta A_s.$$

This partly repeats Remark 5.3.1. If A is continuous, we obtain that $[X] = [M]$, even if X (hence M) is only RCLL.

Now suppose that X is a *continuous* semimartingale. As already pointed out in Remark 5.3.1, the processes M and A can then also be chosen continuous. A simple result from analysis [$\rightarrow$ *exercise*] says that

(1.3) any continuous function of finite variation has zero quadratic variation along any sequence $(\Pi_n)_{n \in \mathbb{N}}$ of partitions of $[0, \infty)$ whose mesh size $|\Pi_n|$ goes to 0 as $n \to \infty$.

(Note that this is a variant of the result already mentioned in Remark 4.1.5 in Chapter 4.) So if the semimartingale X is continuous, then its (unique) finite variation part A has zero quadratic variation, and its (unique) local martingale part M has quadratic variation $[M] = \langle M \rangle$; see Remark 5.1.2 in Chapter 5. The covariation of M and A is thus also zero by Cauchy–Schwarz. A continuous semimartingale X with canonical decomposition $X = X_0 + M + A$ therefore has the quadratic variation $[X] = \langle X \rangle = [M] = \langle M \rangle$ which is again continuous.

Now let us return to the transformation $f(X)$ of a semimartingale X by a function f. In the simplest case, the answer to our basic question in this section looks as follows.

Theorem 1.1 (Itô's formula I). *Suppose $X = (X_t)_{t \geq 0}$ is a continuous real-valued semimartingale and $f : \mathbb{R} \to \mathbb{R}$ is in C^2. Then $f(X) = (f(X_t))_{t \geq 0}$ is again a continuous (real-valued) semimartingale, and we explicitly have P-a.s.*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) \, \mathrm{d}X_s + \frac{1}{2} \int_0^t f''(X_s) \, \mathrm{d}\langle X \rangle_s \tag{1.4}$$

for all $t \geq 0$.

Remarks. 1) Not only the result is important, but also the basic idea for its proof.

2) The $\mathrm{d}X$-integral in (1.4) is a stochastic integral; it is well defined because X is a semimartingale and $f'(X)$ is adapted and continuous, hence predictable and locally bounded. The $\mathrm{d}\langle X \rangle$-integral is a classical Lebesgue–Stieltjes integral because $\langle X \rangle$ has increasing trajectories; it is also well defined because $f''(X)$ is also predictable and locally bounded.

3) In purely formal *differential notation*, (1.4) is usually written more compactly as

$$\mathrm{d}f(X_t) = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)\,\mathrm{d}\langle X\rangle_t = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)\,\mathrm{d}\langle M\rangle_t, \tag{1.5}$$

using that $\langle X\rangle = \langle M\rangle$.

4) Comparing (1.1), (1.2) to (1.5), (1.4) shows that we have in comparison to the classical chain rule an extra *second-order term* coming from the *quadratic variation* of X (or here more precisely from the quadratic variation of the martingale part M of X). This is the important point to remember, and it also shows up in the proof.

5) One can view Itô's formula and its proof as a purely analytical result which provides an *extension of the chain rule* for $f \circ x$ to functions x that have a nonzero quadratic variation. This has been pointed out and developed by Hans Föllmer [8]. Not surprisingly, relaxing the assumptions on x then requires stronger assumptions on f than in the classical case (C^2 instead of C^1).

6) To see the financial relevance of Itô's formula, think of X as some underlying financial asset and of $Y = f(X)$ as a new product obtained from the underlying by a possibly nonlinear transformation f. Then (1.4) or (1.5) show us how the product reacts to changes in the underlying. The important message of Theorem 1.1 is then that when using stochastic models (for X), *a simple linear approximation is not good enough; one must also account for the second-order behaviour of X.* $\diamond$

Proof of Theorem 1.1. The easiest way to remember both the result and its proof for the case where X is continuous is via the following *quick and dirty* argument: "A Taylor expansion at the infinitesimal level gives

$$\mathrm{d}f(X_t) = f(X_t) - f(X_{t-\mathrm{d}t}) = f'(X_t)\,\mathrm{d}X_t + \frac{1}{2}f''(X_t)(\,\mathrm{d}X_t)^2,$$

and $(\mathrm{d}X_t)^2 = (X_t - X_{t-\mathrm{d}t})^2 = \langle X\rangle_t - \langle X\rangle_{t-\mathrm{d}t} = \mathrm{d}\langle X\rangle_t$." Note, however, that this reasoning is purely formal and does not constitute a correct proof. (For example, it does not explain why we stop at the second and not at another higher order in the expansion.)

To make the above idea rigorous, we write for non-infinitesimal increments

$$f(X_{t_{i+1}\wedge t}) - f(X_{t_i}) = f'(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i}) + \frac{1}{2} f''(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i})^2 + R_i,$$

where R_i stands for the error term in the Taylor expansion and the t_i come from a partition Π_n of $[0,\infty)$. Now we sum over the $t_i \le t$ and obtain on the left-hand side a telescoping sum which equals $f(X_t) - f(X_0)$. When we study the terms on the right-hand side, we first recall the convergence

$$Q_t^{\Pi_n} := \sum_{t_i \in \Pi_n,\, t_i \le t} (X_{t_{i+1}\wedge t} - X_{t_i})^2 \longrightarrow \langle X \rangle_t \qquad \text{as } |\Pi_n| \to 0$$

from Theorem 5.1.1; see also Remark 5.3.1. This implies firstly by a weak convergence argument that

$$\frac{1}{2} \sum_{t_i \in \Pi_n,\, t_i \le t} f''(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i})^2 \longrightarrow \frac{1}{2} \int_0^t f''(X_s)\, \mathrm{d}\langle X \rangle_s,$$

and secondly by a careful estimate that

$$\sum_{t_i \in \Pi_n,\, t_i \le t} |R_i| \longrightarrow 0.$$

(This is exactly the point where the mathematical analysis shows why the second order is the correct order of expansion.) As a consequence, the sums

$$\sum_{t_i \in \Pi_n,\, t_i \le t} f'(X_{t_i})(X_{t_{i+1}\wedge t} - X_{t_i})$$

must also converge, and the dominated convergence theorem for stochastic integrals then implies that the limit is $\int_0^t f'(X_s)\, \mathrm{d}X_s$. **q.e.d.**

Example. For $X = W$ a Brownian motion and $f(x) = x^2$, we obtain $f'(x) = 2x$, $f''(x) \equiv 2$ and therefore

$$W_t^2 = W_0^2 + \int_0^t 2W_s\, \mathrm{d}W_s + \frac{1}{2} \int_0^t 2\, \mathrm{d}\langle W \rangle_s.$$

Using $W_0 = 0$ and the fact that BM has quadratic variation $\langle W \rangle_t = t$, hence $\mathrm{d}\langle W \rangle_s = \mathrm{d}s$, gives

$$W_t^2 = 2\int_0^t W_s \,\mathrm{d}W_s + \int_0^t \mathrm{d}s = 2\int_0^t W_s \,\mathrm{d}W_s + t$$

or rewritten

$$\int_0^t W_s \,\mathrm{d}W_s = \frac{1}{2}W_t^2 - \frac{1}{2}t.$$

This ties up with the example we have seen in Section 5.2.

Before moving on to more examples, we need an extension of Theorem 1.1.

Theorem 1.2 (Itô's formula II). *Suppose $X = (X_t)_{t\ge 0}$ is a general $\mathbb{R}^d$-valued semimartingale and $f : \mathbb{R}^d \to \mathbb{R}$ is in C^2. Then $f(X) = (f(X_t))_{t\ge 0}$ is again a (real-valued) semimartingale, and we explicitly have P-a.s. for all $t \ge 0$*

1) *if X has continuous trajectories:*

$$(1.6) \qquad f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x^i}(X_s)\,\mathrm{d}X_s^i + \frac{1}{2}\sum_{i,j=1}^d \int_0^t \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s)\,\mathrm{d}\langle X^i, X^j\rangle_s,$$

or in more compact notation, using subscripts to denote partial derivatives,

$$\mathrm{d}f(X_t) = \sum_{i=1}^d f_{x^i}(X_t)\,\mathrm{d}X_t^i + \frac{1}{2}\sum_{i,j=1}^d f_{x^i x^j}(X_t)\,\mathrm{d}\langle X^i, X^j\rangle_t.$$

2) *if $d = 1$ (so that X is real-valued, but not necessarily continuous):*

$$(1.7) \qquad f(X_t) = f(X_0) + \int_0^t f'(X_{s-})\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_{s-})\,\mathrm{d}[X]_s + \sum_{0<s\le t}\Big(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s - \frac{1}{2}f''(X_{s-})(\Delta X_s)^2\Big).$$

Proof. See Protter [13, Section II.7]. **q.e.d.**

Remark. There is of course also a version of Itô's formula for general $\mathbb{R}^d$-valued semimartingales (which contains both 1) and 2) as special cases). It looks similar to part 2) of Theorem 1.2, but has in addition sums like in part 1), with $\langle\,\cdot\,,\,\cdot\,\rangle$ replaced by $[\,\cdot\,,\,\cdot\,]$. And of course one could also write (1.7) in differential form. ◇

If X is continuous, one frequently useful *simplification* of (1.6) arises if one or several of the components of X are of finite variation. If X^k, say, is of finite variation, then we know from (1.3) that $\langle X^k\rangle \equiv 0$ and hence also $\langle X^i, X^k\rangle \equiv 0$ for all i by Cauchy–Schwarz. (Recall that we have already used such an argument before Theorem 1.1.) This implies that all the second-order terms containing X^k will vanish; hence we do not need all the corresponding partial derivatives, and so we can also relax the assumptions on f in that regard.

Example 1.3. The CRR binomial model can be written as

$$\frac{\widetilde{S}^0_k - \widetilde{S}^0_{k-1}}{\widetilde{S}^0_{k-1}} = r,$$

$$\frac{\widetilde{S}^1_k - \widetilde{S}^1_{k-1}}{\widetilde{S}^1_{k-1}} = Y_k - 1 =: R_k = E[R_k] + (R_k - E[R_k]).$$

Note that the terms in brackets above has expectation 0 and a variance which depends on the distribution of the R_k. Passing from time steps of size 1 to $\mathrm{d}t$ and noting that Brownian increments have expectation 0 like the term $R_k - E[R_k]$, a continuous-time analogue would be of the form

$$\frac{\mathrm{d}\widetilde{S}^0_t}{\widetilde{S}^0_t} = r\,\mathrm{d}t, \tag{1.8}$$

$$\frac{\mathrm{d}\widetilde{S}^1_t}{\widetilde{S}^1_t} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{1.9}$$

(More accurately, we should put $\mathrm{d}\widetilde{S}^0_t/\widetilde{S}^0_{t-}$ and $\mathrm{d}\widetilde{S}^1_t/\widetilde{S}^1_{t-}$. But as both $\widetilde{S}^0$ and $\widetilde{S}^1$ turn out to be continuous, the difference does not matter.)

Of course, the equation (1.8) for $\widetilde{S}^0$ is just a very simple ordinary differential equation (ODE), whose solution for the starting value $\widetilde{S}^0_0 = 1$ is $\widetilde{S}^0_t = e^{rt}$. The equation (1.9) for

$\widetilde{S}^1$ is a *stochastic differential equation (SDE)*, and its solution is given by the *geometric Brownian motion (GBM)*

$$\widetilde{S}^1_t = \widetilde{S}^1_0 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right)t\right) \qquad \text{for } t \geq 0. \tag{1.10}$$

Note the possibly surprising term $-\frac{1}{2}\sigma^2$. To see that this is indeed a solution, we write

$$\widetilde{S}^1_t = f(W_t, t) \qquad \text{with } f(x,t) = \widetilde{S}^1_0\, e^{\sigma x + (\mu - \frac{1}{2}\sigma^2)t}.$$

We now apply Itô's formula (1.6) for $d = 2$ to $X_t = (W_t, t)$. As the second component $X^{(2)}_t = t$ is continuous and increasing, it has finite variation; so (1.6) simplifies and we only need the derivatives

$$\begin{aligned} f_x &= \frac{\partial f}{\partial x} = \sigma f, \\ f_t &= \frac{\partial f}{\partial t} = \left(\mu - \frac{1}{2}\sigma^2\right) f, \\ f_{xx} &= \frac{\partial^2 f}{\partial x^2} = \sigma^2 f. \end{aligned}$$

Then we get, by using that $\langle W \rangle_t = t$ and $f(W_t, t) = \widetilde{S}^1_t$, that

$$\begin{aligned} \mathrm{d}\widetilde{S}^1_t &= f_x(W_t,t)\,\mathrm{d}W_t + f_t(W_t,t)\,\mathrm{d}t + \frac{1}{2} f_{xx}(W_t,t)\,\mathrm{d}\langle W \rangle_t \\ &= \sigma \widetilde{S}^1_t\,\mathrm{d}W_t + \left(\mu - \frac{1}{2}\sigma^2\right)\widetilde{S}^1_t\,\mathrm{d}t + \frac{1}{2}\sigma^2 \widetilde{S}^1_t\,\mathrm{d}t \\ &= \widetilde{S}^1_t(\sigma\,\mathrm{d}W_t + \mu\,\mathrm{d}t), \end{aligned}$$

exactly as claimed. Note that we did not argue (as one should and can) that the above explicit process in (1.10) is the only solution of (1.9).

Example. If $X = (X_t)_{t \geq 0}$ is a continuous real-valued semimartingale null at 0, then

$$Z_t := e^{X_t - \frac{1}{2}\langle X \rangle_t} \qquad \text{for } t \geq 0 \tag{1.11}$$

is the unique solution of the SDE

$$\mathrm{d}Z_t = Z_t\,\mathrm{d}X_t, \qquad Z_0 = 1.$$

Put differently, this means that Z satisfies

$$Z_t = 1 + \int_0^t Z_s \,\mathrm{d}X_s \qquad \text{for all } t \geq 0,\ P\text{-a.s.}$$

Checking that the above Z does satisfy the above SDE, as well as proving uniqueness of the solution, is a good [$\rightarrow$ *exercise*] in the use of Itô's formula.

Definition. For a general real-valued semimartingale X null at 0, the *stochastic exponential* of X is defined as the unique solution Z of the SDE

$$\mathrm{d}Z_t = Z_{t-}\,\mathrm{d}X_t, \qquad Z_0 = 1,$$

i.e.,

$$Z_t = 1 + \int_0^t Z_{s-}\,\mathrm{d}X_s \qquad \text{for all } t \geq 0,\ P\text{-a.s.},$$

and it is denoted by $\mathcal{E}(X) := Z$.

From the preceding example, we have the explicit formula $\mathcal{E}(X) = \exp(X - \frac{1}{2}\langle X\rangle)$ when X is continuous and null at 0. For general X, an explicit formula is given in Protter [13, Theorem II.37]. Note that $Z = \mathcal{E}(X)$ can become 0 or negative when X has jumps; in fact, the properties of jumps of stochastic integrals yield

$$Z_t - Z_{t-} = \Delta Z_t = \Delta\left(1 + \int Z_{s-}\,\mathrm{d}X_s\right) = Z_{t-}\Delta X_t,$$

and this shows that $Z_t = Z_{t-}(1 + \Delta X_t)$ so that $Z = \mathcal{E}(X)$ changes sign between $t-$ and t whenever $1 + \Delta X_t < 0$, i.e. when X has a jump $\Delta X_t < -1$.

Example 1.4. Suppose W is a Brownian motion, $T \in (0, \infty)$ is fixed and $h : I\!R \to I\!R$ is a measurable function with $h(W_T) \in L^1$. Then clearly

$$M_t := E[h(W_T)\,|\,\mathcal{F}_t] \qquad \text{for } 0 \leq t \leq T$$

is a martingale. But writing

$$M_t = E[h(W_t + W_T - W_t)\,|\,\mathcal{F}_t]$$

and using that W_t is $\mathcal{F}_t$-measurable and $W_T - W_t$ is independent of $\mathcal{F}_t$ and $\sim \mathcal{N}(0, T-t)$ shows that we also have

$$M_t = E[h(x + W_T - W_t)]\big|_{x=W_t} = f(W_t, t)$$

with

$$f(x,t) = E[h(x + W_T - W_t)] = \int_{-\infty}^{\infty} h(x+y) \frac{1}{\sqrt{2\pi(T-t)}} e^{-\frac{y^2}{2(T-t)}} \, \mathrm{d}y.$$

So $f(\,\cdot\,, t)$, as a function of x for fixed $t < T$, is the convolution of h with a function in C^∞ and therefore also C^∞ with respect to x, and $f(x, \,\cdot\,)$ is clearly in C^1 with respect to t as long as $t < T$. Therefore Itô's formula may be applied and gives

$$\text{(1.12)} \qquad M_t = M_0 + \int_0^t f_x(W_s, s) \, \mathrm{d}W_s + \int_0^t \left(f_t + \frac{1}{2} f_{xx} \right)(W_s, s) \, \mathrm{d}s \qquad \text{for } 0 \le t < T.$$

Now one can check by laborious analysis that the function $f(x,t)$ satisfies the partial differential equation (PDE) $f_t + \frac{1}{2} f_{xx} = 0$; or one can use the fact that the canonical decomposition of a special semimartingale (like the martingale M) is unique. (Alternatively, one can use that any continuous local martingale of finite variation is constant.) Any of these leads to the conclusion that the ds-integral in (1.12) must vanish identically because it is continuous and adapted, hence predictable, and of finite variation like any ds-integral. By letting $t \nearrow T$ in (1.12), we therefore obtain the representation

$$h(W_T) = M_T = M_0 + \int_0^T f_x(W_s, s) \, \mathrm{d}W_s$$

of the random variable $h(W_T)$ as an initial value M_0 plus a stochastic integral with respect to the Brownian motion W. A more general result in that direction is given in Section 6.3.

Example. An *Itô process* is a stochastic process of the form

$$X_t = X_0 + \int_0^t \mu_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s \qquad \text{for } t \ge 0$$

for some Brownian motion W, where μ and σ are predictable processes satisfying appropriate integrability conditions (e.g. $\int_0^T (|\mu_s| + |\sigma_s|^2) \, \mathrm{d}s < \infty$ P-a.s. for every $T < \infty$).

More generally, X, μ, W could be vector-valued and σ could be matrix-valued, of course all with appropriate dimensions. For any C^2-function f, the process $f(X)$ is then again an Itô process, and Itô's formula gives

$$f(X_t) = f(X_0) + \int_0^t \left(f'(X_s)\mu_s + \frac{1}{2} f''(X_s)\sigma_s^2 \right) \mathrm{d}s + \int_0^t f'(X_s)\sigma_s \, \mathrm{d}W_s.$$

This is another good [$\rightarrow$ *exercise*] for using Itô's formula.

Example. For any two real-valued (RCLL) semimartingales X and Y, the *product rule* is obtained by applying Itô's formula with the function $f(x, y) = xy$. The result says that

$$X_t Y_t = X_0 Y_0 + \int_0^t Y_{s-} \, \mathrm{d}X_s + \int_0^t X_{s-} \, \mathrm{d}Y_s + [X, Y]_t$$

or compactly in differential notation

$$\mathrm{d}(XY) = Y_- \, \mathrm{d}X + X_- \, \mathrm{d}Y + \mathrm{d}[X, Y].$$

If both X and Y are continuous, this yields

$$\mathrm{d}(XY) = Y \, \mathrm{d}X + X \, \mathrm{d}Y + \mathrm{d}\langle X, Y \rangle.$$

Example. Let $W = (W_t)_{t \geq 0}$ be a Brownian motion, $a < 0 < b$ and

$$\tau_{a,b} := \inf \{ t \geq 0 : W_t > b \text{ or } W_t < a \}$$

the first time that BM leaves the interval $[a, b]$ around 0. Then classical results about the *ruin problem for Brownian motion* say that

$$E[\tau_{a,b}] = |a| \, b \qquad (\text{so that } \tau_{a,b} < \infty \ P\text{-a.s.})$$

and

$$P[W_{\tau_{a,b}} = b] = \frac{|a|}{b - a} = 1 - P[W_{\tau_{a,b}} = a]. \tag{1.13}$$

It is also known, or can be computed from (1.13), that $E[W_{\tau_{a,b}}] = 0$.

In order to compute the *covariance* of $\tau_{a,b}$ and $W_{\tau_{a,b}}$, we start with the function $f(x,t) = -\frac{1}{3}x^3 + tx$. Then clearly $f_t + \frac{1}{2}f_{xx} \equiv 0$ so that Itô's formula shows that

$$M_t := f(W_t, t) = 0 + \int_0^t f_x(W_s, s)\,\mathrm{d}W_s$$

is like W a continuous local martingale, and so is then the stopped process $M^{\tau_{a,b}}$. But

$$M_t^{\tau_{a,b}} = M_{t\wedge\tau_{a,b}} = -\frac{1}{3}\big(W_t^{\tau_{a,b}}\big)^3 + (t \wedge \tau_{a,b})W_t^{\tau_{a,b}}$$

is bounded by a constant for $t \leq T$ as $|W^{\tau_{a,b}}| \leq \max(|a|, b)$, and so $M^{\tau_{a,b}}$ is a martingale on $[0,T]$ for each $T < \infty$. This directly implies that

$$0 = E\big[M_0^{\tau_{a,b}}\big] = E\big[M_T^{\tau_{a,b}}\big] = -\frac{1}{3}E\big[W_{\tau_{a,b}\wedge T}^3\big] + E\big[(\tau_{a,b}\wedge T)W_{\tau_{a,b}\wedge T}\big],$$

and letting $T \to \infty$ yields by dominated convergence, also using $\tau_{a,b} \in L^1$, that

$$0 = -\frac{1}{3}E\big[W_{\tau_{a,b}}^3\big] + E\big[\tau_{a,b}W_{\tau_{a,b}}\big].$$

Hence we find

$$\mathrm{Cov}\big(\tau_{a,b}, W_{\tau_{a,b}}\big) = E\big[\tau_{a,b}W_{\tau_{a,b}}\big] = \frac{1}{3}E\big[W_{\tau_{a,b}}^3\big] = \frac{1}{3}|a|b(b-|a|),$$

where the last equality is obtained by computing with the known (two-point) distribution of $W_{\tau_{a,b}}$ given in (1.13).

6.2 Girsanov's theorem

In Section 6.1, we have seen that the family of semimartingales is *invariant* under a transformation by a C^2-function, i.e., $f(X)$ is a semimartingale whenever X is a semimartingale and $f \in C^2$. In this section, our goal is to show that the class of semimartingales is also invariant under a change to an equivalent probability measure.

Suppose we have P and a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$. Assuming that $Q \approx P$ on $\mathcal{F}$ (or $\mathcal{F}_\infty$) can be too restrictive; so we fix $T \in (0, \infty)$ and assume only that $Q \approx P$ on $\mathcal{F}_T$. If we have this for every $T < \infty$, we call Q and P *locally equivalent* and write $Q \overset{\text{loc}}{\approx} P$. For an infinite horizon, this is usually strictly weaker than $Q \approx P$. (Also, one must be careful with the filtration and the usual conditions, but we do not discuss these technical issues.)

To start, fix $T \in (0, \infty)$ for simplicity and suppose that $Q \approx P$ on $\mathcal{F}_T$. Denote by

$$Z_t := Z_t^{Q;P} := E_P\left[\frac{\mathrm{d}Q|_{\mathcal{F}_T}}{\mathrm{d}P|_{\mathcal{F}_T}} \,\middle|\, \mathcal{F}_t\right] \qquad \text{for } 0 \leq t \leq T \tag{2.1}$$

the *density process* of Q with respect to P on $[0, T]$, choosing an RCLL version of this P-martingale on $[0, T]$. Because $Q \approx P$ on $\mathcal{F}_T$, we have $Z > 0$ on $[0, T]$, meaning that $P[Z_t > 0, \forall t \in [0, T]] = 1$, and because Z is a P-(super)martingale, we even have $\inf_{0 \leq t \leq T} Z_t > 0$ P-a.s. by the so-called *minimum principle for supermartingales*; see Dellacherie/Meyer [5, Theorem VI.17]. This implies that also $Z_- > 0$ on $[0, T]$ so that $1/Z_-$ is well defined and adapted and left-continuous, hence also predictable and locally bounded.

In perfect analogy to Lemma 2.3.1, we now have

Lemma 2.1. *Suppose that $Q \approx P$ on $\mathcal{F}_T$ and define $Z = Z^{Q;P}$ as in (2.1). Then:*

1) *For $s \leq t \leq T$ and every U_t which is $\mathcal{F}_t$-measurable and either ≥ 0 or in $L^1(Q)$, we have the* Bayes formula

$$E_Q[U_t \,|\, \mathcal{F}_s] = \frac{1}{Z_s} E_P[Z_t U_t \,|\, \mathcal{F}_s] \qquad Q\text{-a.s.}$$

2) *An adapted process $Y = (Y_t)_{0 \leq t \leq T}$ is a (local) Q-martingale on $[0, T]$ if and only if the product ZY is a (local) P-martingale on $[0, T]$.*

Of course, if $Q \overset{\text{loc}}{\approx} P$, we can use Lemma 2.1 for any $T < \infty$ and hence obtain a statement for processes $Y = (Y_t)_{t \geq 0}$ on $[0, \infty)$. One consequence of part 2) of Lemma 2.1 (with $Y := 1/Z$) is also that $\frac{1}{Z}$ is a Q-martingale, more precisely on $[0, T]$ if $Q \approx P$ on $\mathcal{F}_T$, or even on $[0, \infty)$ if $Q \overset{\text{loc}}{\approx} P$. Furthermore, it is easy to check that $\frac{1}{Z}$ is the density process of P with respect to Q (again on $[0, T]$ or on $[0, \infty)$, respectively).

The next result now proves the announced basic result.

Theorem 2.2 (Girsanov). *Suppose that $Q \overset{\text{loc}}{\approx} P$ with density process Z. If M is a local P-martingale null at* 0*, then*

$$\widetilde{M} := M - \int \frac{1}{Z} \, \mathrm{d}[Z, M]$$

is a local Q-martingale null at 0*. In particular, every P-semimartingale is also a Q-semimartingale (and vice versa, by symmetry).*

Proof. The second assertion is very easy to prove from the first; we simply write

$$X = X_0 + M + A = X_0 + \widetilde{M} + \left(A + \int \frac{1}{Z} \, \mathrm{d}[Z, M] \right) = X_0 + \widetilde{M} + \widetilde{A}$$

and observe that $\widetilde{A} := A + \int \frac{1}{Z} \, \mathrm{d}[Z, M]$ is of finite variation. Note that $\int \frac{1}{Z} \, \mathrm{d}[Z, M]$ is defined pathwise because $[Z, M]$ is of finite variation; so this requires no stochastic integration, nor predictability of the integrand.

For proving the first assertion, note that the definition of the optional covariation process implies that the difference $ZM - [Z, M]$ is a local P-martingale like M and Z. (To argue this in an alternative manner, we could use the product rule which gives $ZM - [Z, M] = \int Z_- \, \mathrm{d}M + \int M_- \, \mathrm{d}Z$, which is a local P-martingale like M and Z.) So by Lemma 2.1,

$$M - \frac{1}{Z}[Z, M] \text{ is a local } Q\text{-martingale.}$$

Using the product rule gives

$$\frac{1}{Z}[Z,M] = \int [Z,M]_- \,\mathrm{d}\left(\frac{1}{Z}\right) + \int \frac{1}{Z_-}\,\mathrm{d}[Z,M] + \left[\frac{1}{Z},[Z,M]\right]. \tag{2.2}$$

Because $[Z,M]$ is of finite variation, the last term equals

$$\left[\frac{1}{Z},[Z,M]\right] = \sum \Delta\left(\frac{1}{Z}\right)\Delta[Z,M] = \int \Delta\left(\frac{1}{Z}\right)\mathrm{d}[Z,M]$$

so that the last two terms in (2.2) add up to $\int \frac{1}{Z}\,\mathrm{d}[Z,M]$. Because $\frac{1}{Z}$ is a local Q-martingale, so is the stochastic integral $\int [Z,M]_-\,\mathrm{d}\left(\frac{1}{Z}\right)$ because its integrand is locally bounded. So we obtain

$$\widetilde{M} = M - \int \frac{1}{Z}\,\mathrm{d}[Z,M] = \left(M - \frac{1}{Z}[Z,M]\right) - \int [Z,M]_-\,\mathrm{d}\left(\frac{1}{Z}\right),$$

and we see that this is a local Q-martingale. **q.e.d.**

In many situations, it is more convenient to do computations not in terms of Z, but rather with its so-called stochastic logarithm. Suppose in general that Y is a semimartingale with $Y_- > 0$ (on $[0,T]$ or $[0,\infty)$, respectively). Then we can define a semimartingale null at 0 by $L := \int \frac{1}{Y_-}\,\mathrm{d}Y$, we have $\mathrm{d}Y = Y_-\,\mathrm{d}L$ by construction, and so we obtain

$$Y = Y_0\,\mathcal{E}(L) > 0 \qquad \text{with a semimartingale } L \text{ null at } 0.$$

It is also clear that L is continuous if and only if Y is continuous, and that L is a local P-martingale if and only if Y is a local P-martingale. This L is called the *stochastic logarithm* of Y. Note that because of the quadratic variation, we do not have $L = \log Y$, not even if Y is continuous; see the explicit formula (1.11) in Section 6.1.

In the situation here, Z is a P-martingale > 0, hence has $Z_- > 0$ as discussed above, and so applying the above with $Y := Z$ yields $Z = Z_0\,\mathcal{E}(L)$, where L is like Z a local P-martingale.

Theorem 2.3 (Girsanov, continuous version). *Suppose that* $Q \overset{\text{loc}}{\approx} P$ *with a density process* Z *which is* continuous. *Write* $Z = Z_0\,\mathcal{E}(L)$. *If* M *is a local* P*-martingale null at*

0, then

$$\widetilde{M} := M - [L, M] = M - \langle L, M \rangle$$

is a local Q-martingale null at 0.

More specifically, if W is a P-Brownian motion, then $\widetilde{W}$ is a Q-Brownian motion. In particular, if $L = \int \nu \,\mathrm{d}W$ for some $\nu \in L^2_{\mathrm{loc}}(W)$, then

$$\widetilde{W} = W - \left\langle \int \nu \,\mathrm{d}W, W \right\rangle = W - \int \nu_s \,\mathrm{d}s$$

so that the P-Brownian motion $W = \widetilde{W} + \int \nu_s \,\mathrm{d}s$ becomes under Q a Brownian motion with (instantaneous) drift ν.

Proof. Because $Z = Z_0 \,\mathcal{E}(L)$ satisfies $\mathrm{d}Z = Z_- \,\mathrm{d}L$, we have $[Z, M] = \int Z_- \,\mathrm{d}[L, M]$ and hence $\int \frac{1}{Z} \,\mathrm{d}[Z, M] = \int \frac{Z_-}{Z} \,\mathrm{d}[L, M] = [L, M]$ by continuity of Z. So the first assertion follows directly from Theorem 2.2, and $[L, M] = \langle L, M \rangle$ because L is continuous like Z.

The assertion for $\widetilde{W}$ needs some extra work as it relies on the so-called *Lévy characterisation of Brownian motion* that we have not discussed here. **q.e.d.**

In all the above discussions, we have assumed that Q is already given and have then studied its effect on given processes. But in mathematical finance, we often want to proceed the other way round: We start with a process $S = (S_t)_{0 \le t \le T}$ of discounted asset prices and want to find or construct some $Q \approx P$ on $\mathcal{F}_T$ such that S becomes a local Q-martingale. Let us now see how we can tackle this problem by reverse-engineering the preceding theory. We begin very generally and successively become more specific. Moreover, the goal here is not to remember a specific result, but rather to understand how to approach the problem in a systematic way.

We start with a local P-martingale L null at 0 and define $Z := \mathcal{E}(L)$ so that Z is like L a local P-martingale, with $Z_0 = 1$. If we also have $\Delta L > -1$ (and this holds of course in particular if L is continuous), then we have in addition $Z > 0$. This uses that $\Delta Z = Z_- \Delta L$ so that $Z = Z_-(1 + \Delta L)$, which implies that Z never changes sign as long as $\Delta L > -1$.

Suppose now that Z is a *true* P-martingale on $[0,T]$; this amounts to imposing suitable extra conditions on L. Then we can define a probability measure $Q \approx P$ on $\mathcal{F}_T$ by setting $\mathrm{d}Q := Z_T \, \mathrm{d}P$, and the density process of Q with respect to P on $[0,T]$ is then by construction the P-martingale Z. In particular, if L is continuous, also Z is continuous.

In a bit more detail, $Z = \mathcal{E}(L)$ is in the present situation a local P-martingale > 0 on $[0,T]$ and therefore a P-supermartingale starting at 1. So $t \mapsto E[Z_t]$ is decreasing, and one can easily check that Z is a P-martingale on $[0,T]$ if and only if $t \mapsto E[Z_t]$ is identically 1 on $[0,T]$, or also if and only if $E[Z_T] = 1$. However, expressing this directly in terms of L is more tricky, and one has only sufficient conditions on L that ensure $E[\mathcal{E}(L)_T] = 1$. The most famous of these is the *Novikov condition*: If L is a continuous local martingale null at 0 and $E[e^{\frac{1}{2}\langle L \rangle_T}] < \infty$, then $Z = \mathcal{E}(L)$ is a martingale on $[0,T]$.

Now start with an $\mathbb{R}^d$-valued process $S = (S_t)_{0 \le t \le T}$ and suppose that S is a P-semimartingale. For each i, the coordinate S^i can then (in general non-uniquely) be written as

$$S^i = S^i_0 + M^i + A^i$$

with a local P-martingale M^i and an adapted process A^i of finite variation, both null at 0. By Theorem 2.2,

$$\widetilde{M}^i = M^i - \int \frac{1}{Z} \, \mathrm{d}[Z, M^i]$$

is then a local Q-martingale, and of course we have

$$S^i = S^i_0 + \widetilde{M}^i + \left(A^i + \int \frac{1}{Z} \, \mathrm{d}[Z, M^i] \right) = S^i_0 + \widetilde{M}^i + \widetilde{A}^i.$$

So S^i is a local Q-martingale (or, equivalently, Q is an ELMM for S^i) if and only if

$$\widetilde{A}^i = A^i + \int \frac{1}{Z} \, \mathrm{d}[Z, M^i] \quad \text{is a local } Q\text{-martingale.}$$

One sufficient condition is obviously that

$$A^i + \int \frac{1}{Z} \, \mathrm{d}[Z, M^i] \equiv 0. \tag{2.3}$$

This should be viewed as a condition on Z or, equivalently, on L. In general, because $\mathrm{d}Z = Z_- \, \mathrm{d}L$, we have

$$[Z, M^i] = \int Z_- \, \mathrm{d}[L, M^i]$$

and $\Delta Z = Z_- \Delta L$, hence

$$Z = Z_- + \Delta Z = Z_-(1 + \Delta L)$$

and so

$$\frac{Z_-}{Z} = \frac{1}{1 + \Delta L}.$$

So in terms of L, the sufficient condition (2.3) can be written as

$$A^i + \int \frac{1}{1 + \Delta L} \, \mathrm{d}[L, M^i] \equiv 0.$$

If L is continuous, this simplifies further to

$$A^i + \langle L, M^i \rangle \equiv 0;$$

this could alternatively also be derived directly from Theorem 2.3. As a condition on L in terms of M and A, this is fairly explicit. Note that this is actually a system of d conditions (one for each S^i) imposed on a single process L.

In Chapter 7, we shall see how the above ideas can be used to construct explicitly an equivalent martingale measure in the Black–Scholes model of geometric Brownian motion for S. But before that, we study in the next section how local martingales L can (or must) look if we impose more structure on the underlying filtration $\mathbb{F}$.

Remark. Instead of using Theorem 2.2, we could also argue more directly. Suppose again that $Z = \mathcal{E}(L)$ is a true P-martingale > 0 on $[0, T]$, and define $Q \approx P$ on $\mathcal{F}_T$ by $\mathrm{d}Q := Z_T \, \mathrm{d}P$. By Lemma 2.1, S is then a local Q-martingale if and only if ZS is a local P-martingale, and therefore we compute, using the product rule and $\mathrm{d}Z = Z_- \, \mathrm{d}L$,

$$\mathrm{d}(ZS^i) = S^i_- \, \mathrm{d}Z + Z_- \, \mathrm{d}S^i + \mathrm{d}[Z, S^i] = S^i_- \, \mathrm{d}Z + Z_- \, \mathrm{d}M^i + Z_-(\mathrm{d}A^i + \mathrm{d}[L, S^i]).$$

Because both Z and M^i, and hence also their stochastic integrals above, are local P-martingales, we see that Q is an ELMM for S^i if and only if $A^i + [L, S^i]$ is a local P-martingale. A sufficient condition for this is that

$$A^i + [L, S^i] \equiv 0.$$

If L is continuous or if S^i is continuous, this again simplifies to

$$A^i + \langle L, M^i \rangle \equiv 0,$$

because then $[L, A^i] = \sum \Delta L \, \Delta A^i \equiv 0.$ $\diamond$

6.3 Itô's representation theorem

Our goal in this section is to describe all martingales that can exist in a filtration $I\!F$ under the assumption that $I\!F$ is *generated by a Brownian motion* W. This deep structural result goes back to Kiyosi Itô and is the mathematical explanation for the completeness of the Black–Scholes model that we shall see in the next chapter.

We start with a Brownian motion $W = (W_t)_{t\geq 0}$ in $I\!R^m$ defined on a probability space $(\Omega, \mathcal{F}, P)$ without an a priori filtration. We define

$$\mathcal{F}_t^0 := \sigma(W_s, s \leq t) \qquad \text{for } t \geq 0,$$
$$\mathcal{F}_\infty^0 := \sigma(W_s, s \geq 0),$$

and construct the filtration $I\!F^W = (\mathcal{F}_t^W)_{0\leq t\leq\infty}$ by adding to each $\mathcal{F}_t^0$ the class $\mathcal{N}$ of all subsets of P-nullsets in $\mathcal{F}_\infty^0$ to obtain $\mathcal{F}_t^W = \mathcal{F}_t^0 \vee \mathcal{N}$. This so-called *$P$-augmented filtration* $I\!F^W$ is then P-complete (in $(\Omega, \mathcal{F}_\infty^0, P)$, to be accurate) by construction, and one can show, by using the strong Markov property of Brownian motion, that $I\!F^W$ is also automatically right-continuous (so that it satisfies the usual conditions). We usually call $I\!F^W$, slightly misleadingly, the filtration generated by W. One can show that W is also a Brownian motion with respect to $I\!F^W$; the key point is to argue that $W_t - W_s$ is still independent of $\mathcal{F}_s^W \supseteq \mathcal{F}_s^0$, even though $\mathcal{F}_s^W$ contains some sets from $\mathcal{F}_\infty^0$. If one works on $[0, T]$, one replaces ∞ by T; then $\mathcal{F}_\infty^0$ is not needed separately because we use the P-nullsets from the "last" σ-field $\mathcal{F}_T^0$.

Theorem 3.1 (**Itô's representation theorem**). *Suppose that $W = (W_t)_{t\geq 0}$ is a Brownian motion in $I\!R^m$. Then every random variable $H \in L^1(\mathcal{F}_\infty^W, P)$ has a unique representation as*

$$H = E[H] + \int_0^\infty \psi_s \, \mathrm{d}W_s \qquad P\text{-a.s.}$$

for an $I\!R^m$-valued integrand $\psi \in L^2_{\mathrm{loc}}(W)$ with the additional property that $\int \psi \, \mathrm{d}W$ is a $(P, I\!F^W)$-martingale on the closed *interval $[0, \infty]$ (and therefore uniformly integrable).*

Remark. The assumptions on H say that H is integrable and $\mathcal{F}_\infty^W$-measurable. The latter means intuitively that $H(\omega)$ can depend in a measurable way on the entire trajectory $W_{\bullet}(\omega)$ of Brownian motion, but not on any other source of randomness. $\diamond$

Corollary 3.2. *Suppose the filtration $\mathbb{F} = \mathbb{F}^W$ is generated by a Brownian motion W in $\mathbb{R}^m$. Then:*

1) *Every (real-valued) local $(P, \mathbb{F}^W)$-martingale L is of the form $L = L_0 + \int \gamma \, \mathrm{d}W$ for some $\mathbb{R}^m$-valued process $\gamma \in L^2_{\mathrm{loc}}(W)$.*

2) *Every local $(P, \mathbb{F}^W)$-martingale is continuous.*

Proof. For a localizing sequence $(\tau_k)_{k\in\mathbb{N}}$, each $(L - L_0)^{\tau_k}$ is a uniformly integrable martingale N^k, say, and therefore of the form

$$N^k_t = E[N^k_\infty \,|\, \mathcal{F}^W_t] \qquad \text{for } 0 \leq t \leq \infty,$$

for some $N^k_\infty \in L^1(\mathcal{F}^W_\infty, P)$. So Theorem 3.1 and the martingale property of $\int \psi^k \, \mathrm{d}W$ give that $N^k = \int \psi^k \, \mathrm{d}W$ (note that $N^k_0 = 0$). In particular, $N^k = (L - L_0)^{\tau_k}$ is continuous, which means that L is continuous on $[\![0, \tau_k]\!]$. As $\tau_k \nearrow \infty$, L is continuous, and γ is obtained by piecing together the ψ^k via $\gamma := \psi^k$ on $[\![0, \tau_k]\!]$. **q.e.d.**

While the above results are remarkable, the next result is bizarre. Note that in its formulation, the filtration $\mathbb{F}$ is even allowed to be general; but of course we could also take $\mathbb{F} = \mathbb{F}^W$.

Theorem 3.3 (Dudley). *Suppose $W = (W_t)_{t\geq 0}$ is a Brownian motion with respect to P and $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. As usual, set*

$$\mathcal{F}_\infty := \bigvee_{t\geq 0} \mathcal{F}_t = \sigma\Big(\bigcup_{t\geq 0} \mathcal{F}_t\Big).$$

Then every $\mathcal{F}_\infty$-measurable random variable H with $|H| < \infty$ P-a.s. (for example every $H \in L^1(\mathcal{F}_\infty, P)$) can be written as

$$H = \int_0^\infty \psi_s \, \mathrm{d}W_s \qquad P\text{-a.s.}$$

for some integrand $\psi \in L^2_{\mathrm{loc}}(W)$.

Note that there is no constant in the representation of H in Theorem 3.3. Note also that we could for instance take for H a constant and represent this as a stochastic integral of Brownian motion. This makes it almost immediately clear that the integrand ψ in Theorem 3.3 cannot be nice. In fact:

1) In Theorem 3.3, the stochastic integral process $\int \psi \, \mathrm{d}W$ is of course a *local martingale*, and can even be a martingale on $[0, \infty)$, but it is in general *not a martingale on* $[0, \infty]$; if it were, it would have constant expectation 0 up to $+\infty$, which would imply that $E[H] = 0$.

2) In Theorem 3.3, the representation by ψ is *not unique*. In fact, one can easily construct some bounded predictable $\bar{\psi}$ with $0 < \int_0^\infty \bar{\psi}_s^2 \, \mathrm{d}s < \infty$ P-a.s. (so that $\bar{\psi} \not\equiv 0$ and $\bar{\psi} \in L^2_{\mathrm{loc}}(W)$), but nevertheless $\int_0^\infty \bar{\psi}_s \, \mathrm{d}W_s = 0$ P-a.s. Of course, ψ and $\psi + \bar{\psi}$ then represent the same H, but they are different in a nontrivial way.
[*Exercise:* Try to find such a $\bar{\psi}$ — it is not very difficult.]

3) In terms of finance, the integrands ψ appearing in Theorem 3.3 are not nice at all. For one thing, $\int \psi \, \mathrm{d}W$ cannot be bounded from below in general. Indeed, if it were, then $\int \psi \, \mathrm{d}W$ would be a local martingale uniformly bounded from below, hence a supermartingale, and this would imply that we must have $E[H] \le 0$. Moreover, the representation $1 = \int_0^\infty \psi_s \, \mathrm{d}W_s$ looks suspiciously like creating the riskless payoff 1 out of zero initial capital with a self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \psi)$, which would be arbitrage. (But of course, that φ is not admissible, as we have just argued.)

Remark. It is not important for the above results that we work on the infinite interval $[0, \infty]$ or $[0, \infty)$; everything could be done equally well on $[0, T]$ for any $T \in (0, \infty)$. ◇