Lecture Notes on Continuous Time Stochastic Finance

Dr. Jana Bielagk

Last update: 22/07/2018

Contents

1 Essentials on Stochastic Processes in Continuous Time

1.1 Stochastic Processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A real-valued stochastic process $X = (X_t)_{t\geq 0}$ is a family of random variables taking values in $\mathbb{R}$. For each sample point $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called the *sample path* of the process X associated with ω.
Two processes X and Y on the same probability space are called *modifications* of each other if $\mathbb{P}(X_t = Y_t) = 1$ for every $t \geq 0$. They are called *indistinguishable* if almost all sample paths are identical, i.e. if $\mathbb{P}(X_t = Y_t \text{ for all } t \geq 0) = 1$.

Lemma 1.1.

a) If X and Y are indistinguishable, then they are modifications of each other.

b) If X and Y are modifications of each other and if their trajectories are $\mathbb{P}$-a.s. right continuous, then X and Y are indistinguishable.

Proof. The first part is clear. For the second part, see Lemma 21.5(ii) in [Kle08]. □

Example 1.2.
We want to show that two processes can be modifications of each other without being indistinguishable.
Consider $Z \sim \mathcal{N}(0,1)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ and for $t \geq 0$ let $X_t = 0$ and

$$Y_t := \begin{cases} 0, & t \neq Z, \\ 1, & t = Z. \end{cases}$$

On the one hand, $\mathbb{P}(X_t = Y_t) = \mathbb{P}(Z \neq t) = 1$ for all $t \geq 0$, hence X and Y are modifications of each other. On the other hand, the paths of X are continuous, whereas the paths of Y are discontinuous whenever $Z(\omega) \geq 0$, hence X and Y are not indistinguishable.

A trajectory is called *càdlàg* (continue à droite, limite à gauche) if it is right continuous and if limits from the left exist in all points.

Remark 1.3.
Recall that a probability space is called complete *if for each $A \subset \Omega$ with $A \subset B$ where $\mathbb{P}(B) = 0$ one has $A \in \mathcal{F}$. In other words: $\mathcal{F}$ contains all subsets of $\mathbb{P}$-nullsets.*

On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ there shall be a filtration $(\mathcal{F}_t)_{t\geq 0}$, i.e. an increasing sequence of σ-algebras

$$\bigcup_{s<t} \mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$$

which shall satisfy the *usual conditions:*

- $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$ for all $t \geq 0$, i.e. $(\mathcal{F}_t)_{t\geq 0}$ is right continuous,
- $\mathcal{F}_0$ contains all $\mathbb{P}$-nullsets and $\mathcal{F}$ is complete, i.e. the filtration is complete.

Definition 1.4.
Let $X = (X_t)_{t\geq 0}$ be a stochastic process on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

(i) X is (product-)measurable *if $(t,\omega) \mapsto X_t(\omega)$ is $(\mathcal{B}([0,\infty)) \otimes \mathcal{F})$-measurable.*

(ii) X is progressively measurable *if for any $t \geq 0$, the restriction $X|_{[0,t]}$ is $(\mathcal{B}([0,t]) \otimes \mathcal{F}_t)$-measurable.*

(iii) X is adapted *if X_t is $\mathcal{F}_t$-measurable for every $t \geq 0$.*

In particular, if X is progressively measurable, it is already adapted and measurable.

Lemma 1.5. *An adapted process whose sample paths are almost surely either all right- or all left-continuous is progressively measurable.*

Usually we work with right-continuous processes, so there will be no need to distinguish between measurability and progressive measurability. If that is not the case, we still have the following result:

Theorem 1.6. *Let X be measurable and adapted to $(\mathcal{F}_t)_{t\geq 0}$. Then X has a progressively measurable modification.*

Definition 1.7. *Let $X = (X_t)_{t\geq 0}$ be an adapted stochastic process on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Furthermore assume that X is integrable, i.e. $X_t \in L^1(\mathbb{P})$.*

(i) X is called a submartingale *if $\mathbb{E}[X_t|\mathcal{F}_s] \geq X_s$ for all $t \geq s$.*

(ii) X is called a supermartingale *if $\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s$ for all $t \geq s$.*

(iii) X is called a martingale *if $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ for all $t \geq s$.*

Definition 1.8. *An adapted process X on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ is called a* Markov process *if for all $t, s > 0$ and every bounded function $f\colon \mathbb{R} \to \mathbb{R}$ we have*

$$\mathbb{E}[f(X_{t+s})|\mathcal{F}_s] = \mathbb{E}[f(X_{t+s})|\sigma(X_s)].$$

1.2 Brownian Motion

Definition 1.9. *A stochastic process $W = (W_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a standard one-dimensional Brownian motion (BM) if the following conditions hold:*

(i) $W_0 = 0$ $\mathbb{P}$-a.s.;

(ii) W has independent increments, i.e., for all $t, s \geq 0$, $W_{t+s} - W_s$ is independent of $(W_u)_{0\leq u\leq s}$;

(iii) The increments are stationary and normally distributed:

$$W_{t+s} - W_s \sim \mathcal{N}(0, t);$$

(iv) W has almost surely continuous sample paths.

Theorem 1.10. *A standard Brownian motion exists.*

Proof. See Chapter 2 in [KS88] for the construction of a BM or see the lecture on Stochastic Analysis. □

Lemma 1.11. *If W is a standard BM, then $W_t \sim \mathcal{N}(0,t)$ and $\mathrm{Cov}(W_t, W_s) = \min\{s,t\}$.*

Exercise 1. *Prove Lemma 1.11.*

If W is a standard BM (also called standard Wiener process), then $\mathcal{F}_t := \sigma(W_s, s \leq t)$ is its so-called canonical filtration. Its augmented filtration satisfies the usual conditions.

Theorem 1.12. *Let W be a standard BM. Then $(W_t)_{t\geq 0}$ and $(W_t^2 - t)_{t\geq 0}$ are martingales with respect to the canonical filtration.*

Proof. W is clearly adapted to its canonical filtration (by construction of the latter). Let $t > s \geq 0$. Then

$$\mathbb{E}[W_t|\mathcal{F}_s] = \mathbb{E}[W_t - W_s + W_s|\mathcal{F}_s] = \mathbb{E}[W_t - W_s|\mathcal{F}_s] + W_s \overset{(*)}{=} W_s,$$

where we used properties (ii) and (iii) of the BM and known properties of conditional expectation. As $W_t \sim \mathcal{N}(0,t)$, W is also integrable[1], completing the proof that W is indeed a martingale. □

Exercise 2. *Prove the martingale property of $(W_t^2 - t)_{t\geq 0}$.*

Theorem 1.13 (Lévy)**.** *A continuous real-valued process $(X_t)_{t\geq 0}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $X_0 = 0$ is a BM if and only if both $(X_t)_{t\geq 0}$ and $(X_t^2 - t)_{t\geq 0}$ are martingales (with respect to the canonical filtration).*

Remark 1.14. *We call a process continuous if its trajectories are $\mathbb{P}$-a.s. continuous, i.e. if for $\mathbb{P}$-almost all $\omega \in \Omega$, $t \mapsto X_t(\omega)$ is continuous.*

Exercise 3. *Show that for $c > 0$, $V_t := \frac{1}{c} W_{c^2 t} (t \geq 0)$ is a standard BM if W is one.*

Theorem 1.15. *Let W be a standard BM. Then for all $t > s \geq 0$ and every bounded function $f \colon \mathbb{R} \to \mathbb{R}$ we have the Markov property*

$$\mathbb{E}[f(W_t)|(W_u)_{u\leq s}] = \mathbb{E}[f(W_t)|W_s].$$

Proof. Recall that $M_X(u) := \mathbb{E}[e^{uX}]$ is the moment-generating function of X and that the moment-generating function of an $\mathcal{N}(\mu, \sigma^2)$-distributed random variable X is $M_X(u) = e^{\mu u + \frac{\sigma^2 u^2}{2}}$. We use the independent increments property of BM and the fact that $W_{s+t} - W_s \sim \mathcal{N}(0,t)$ in order to do the following calculation:

$$\begin{aligned}
\mathbb{E}[e^{uW_{t+s}}|\mathcal{F}_s] &= e^{uW_s}\mathbb{E}[e^{u(W_{t+s}-W_s)}|\mathcal{F}_s] \\
&= e^{uW_s}\mathbb{E}[e^{u(W_{t+s}-W_s)}] \\
&= e^{uW_s} e^{u^2 \frac{t}{2}} \\
&= e^{uW_s}\mathbb{E}[e^{u(W_{t+s}-W_s)}|W_s] \\
&= \mathbb{E}[e^{uW_{t+s}}|W_s].
\end{aligned}$$

Hence the conditional distribution of W_{t+s} given $\mathcal{F}_s$ is the same of that given W_s, which implies the Markov property. □

[1] $\mathbb{E}[|W_t|]^2 \leq \mathbb{E}[|W_t|^2] = |W_0|^2 + t$.

1.2.1 Excursion: Variation and Quadratic Variation

Definition 1.16. *Let* $[0,T] \subset \mathbb{R}$ *be a finite interval. A set of points* $\tau := \{t_0, t_1, \dots, t_n\}$ *with* $0 = t_0 < t_1 < \dots < t_n = T$ *is called a* partition *of* $[0,T]$ *with mesh* $\|\tau\| := \max\{\,|t_i - t_{i-1}|\mid 1 \le i \le n\}$.

Definition 1.17. *Let* $f\colon [0,T] \to \mathbb{R}$ *be a (deterministic) function. The (total) variation of* f *(on* $[0,T]$*) is defined as*

$$V(f) := \sup\left\{ \sum_{t_i \in \tau} |f(t_i) - f(t_{i-1})| \;\middle|\; \tau \textit{ is a partition of } [0,T] \right\}.$$

If $V(f) < \infty$*, then we say that* f *is of finite variation.*

Definition 1.18. *Let* $f\colon [0,T] \to \mathbb{R}$ *and let* $(\tau_n)_{n\in\mathbb{N}}$ *be a sequence of partitions of* $[0,T]$ *such that* $\|\tau_n\| \to 0$ *as* $n \to \infty$*. The quadratic variation of* f *on* $[0,t] \subset [0,T]$ *along* τ_n *is defined as*

$$V_t^2(f, \tau_n) := \sum_{t_i \in \tau_n \cup \{t\}, t_i \le t} (f(t_i) - f(t_{i-1}))^2.$$

If

$$\langle f \rangle_t := \lim_{n\to\infty} V_t^2(f, \tau_n)$$

exists for all $t \in [0,T]$ *and if this limit is independent of the choice of the sequence of partitions, then the function* $t \mapsto \langle f \rangle_t$ *is called quadratic variation of* f*. If* $t \mapsto \langle f \rangle_t$ *is continuous, then we say that* f *admits the continuous quadratic variation* $\langle f \rangle$.

Theorem 1.19. *If* $f\colon [0,T] \to \mathbb{R}$ *is continuous and of finite variation, then its quadratic variation is zero.*

Theorem 1.20. *Assume that* $f, g\colon [0,T] \to \mathbb{R}$ *are continuous functions. Assume further that* f *admits a continuous quadratic variation* $\langle f \rangle$ *and that* g *has finite (total) variation. Then* $f + g$ *is of continuous quadratic variation* $\langle f + g \rangle = \langle f \rangle$.

Now let's look at sample paths of Brownian motion:

Theorem 1.21.

a) *Sample paths of BM are of infinite (total) variation.*

b) *There exists a sequence of partitions* $(\tau_n)_{n\in\mathbb{N}}$ *of* $[0,T]$ *with* $\lim_{n\to\infty} \|\tau_n\| = 0$ *such that almost surely* $V_t^2(W_\cdot(\omega); \tau_n) \to t$ *for every* $t \in [0,T]$ *as* $n \to \infty$.

For further calculations we have to bear in mind that $\langle W \rangle_t = t$ for any $t \ge 0$.

Remark 1.22. *A suitable partition for b) in Theorem 1.21 is given by*

$$\tau_n = \left\{ t_i^n = Ti2^{-n} \;\middle|\; i \in \{0, 1, \dots, 2^n\} \right\}$$

for $n \in \mathbb{N}$*. For proofs for the above statements, please attend the lecture Stochastic Analysis or consult the literature. For properties of sample paths of Brownian motion, see Section 2.9 in [KS88]. See also Chapters 3 – 5 in [SKY01] or properties of BM.*

2 Stochastic Calculus for Brownian Motion

At the end of last semester we have seen how to calculate integrals of smooth functions of BM with respect to BM in terms of Riemann-Stieltjes integration. When more general integrands are involved (measurable w.r.t. the canonical filtration generated by BM), a different kind of integral is needed, known as Itô integral.
Literature for this chapter: Chapters 3 and 4 in [Øk03]; Chapter 3 in [KS88].
For stochastic integration with respect to other processes than Brownian motion we refer to the parallel lecture on stochastic analysis.

2.1 The Itô Integral

In the last semester, we have seen how to define $\int_0^T f(t, X_t)\mathrm{d}X_t$ for a stochastic process X in a pathwise sense, provided f and X are nice enough. Now we will define $\int_0^T f(t,\omega)\mathrm{d}W_t(\omega)$ in a probabilistic manner by approximating f by elementary functions. We will therefore have to introduce a suitable space of integrable functions.

From now on we always assume to work on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, where the filtration is generated by a standard BM W (and augmented by the null sets).

Definition 2.1. *Let* $\mathcal{V} = \mathcal{V}(0,T)$ *be the class of all functions* $f\colon [0,\infty)\times\Omega \to \mathbb{R}$ *such that*

i) the map $(t,\omega) \to f(t,\omega)$ *is progressively measurable;*

ii) the map f *is square integrable with respect to* $\mathbb{P}\otimes\lambda$*, i.e.,*

$$\mathbb{E}\left[\int_0^T f^2(t,\omega)\mathrm{d}t\right] < \infty.$$

Exercise 4. *Consider the following functions:*

$$\phi_1(t,\omega) := \sum_i W_{t_i}(\omega)\mathbb{1}_{[t_i,t_{i+1})}(t),$$
$$\phi_2(t,\omega) := \sum_i W_{t_{i+1}}(\omega)\mathbb{1}_{[t_i,t_{i+1})}(t).$$

Verify that $\phi_1 \in \mathcal{V}$*, whereas* $\phi_2 \notin \mathcal{V}$*.*

2.1.1 Construction of the stochastic integral

As a first step consider elementary functions, i.e., functions of the form

$$\phi(t,\omega) = \sum_i e_i(\omega)\mathbb{1}_{[t_i,t_{i+1})}(t), \tag{2.1}$$

where $0 \leq t_0 < t_1 < \ldots < t_n \leq T$ and e_i is $\mathcal{F}_{t_i}$-measurable such that $\mathbb{E}[e_i^2] < \infty$ $(i \in \{0,1,\ldots,n-1\})$. For such a function we define (the random variable)

$$I[\phi] := \int_0^T \phi(t,\cdot)\mathrm{d}W_t := \sum_i e_i(W_{t_{i+1}} - W_{t_i}).$$

For the construction of the Itô integral we need an important result stating that the equality of L^2-norms:

$$\|I[\phi]\|_{L^2(\mathbb{P})} = \|\phi\|_{L^2(\mathbb{P}\otimes\lambda)}\,.$$

Lemma 2.2 (Itô isometry)**.** *Let $\phi \in \mathcal{V}$ be a bounded elementary function of the form* (2.1). *Then*

$$\mathbb{E}\left[\left(\int_0^T \phi(t,\cdot)\mathrm{d}W_t\right)^2\right] = \mathbb{E}\left[\int_0^T \phi^2(t,\cdot)\mathrm{d}t\right].$$

Proof. Let $\Delta W_j := W_{t_{j+1}} - W_{t_j}$. Fix indices $i,j \in \mathbb{N}$ with $i \leq j$. Recall that e_i and e_j are measurable w.r.t. $\mathcal{F}_{t_i}$ and $\mathcal{F}_{t_j}$, respectively. Therefore, $e_ie_j\Delta W_i$ and ΔW_j are independent for $i < j$ and e_i is independent of ΔW_i. Consequently, we have that

$$\mathbb{E}[e_ie_j\Delta W_i\Delta W_j] = \mathbb{E}[e_ie_j\Delta W_i]\underbrace{\mathbb{E}[\Delta W_j]}_{=0} = 0$$

and (for $i = j$)

$$\mathbb{E}[e_i^2(\Delta W_i)^2] = \mathbb{E}[e_i^2]\mathbb{E}[(\Delta W_i)^2]] = \mathbb{E}[e_i^2](t_{i+1} - t_i),$$

where we used that $\Delta W_i = W_{t_{i+1}} - W_{t_i} \sim \mathcal{N}(0, t_{i+1} - t_i)$. Combining both results, we have that

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T \phi(t,\cdot)\mathrm{d}W_t\right)^2\right] &= \sum_{i,j}\mathbb{E}[e_ie_j\Delta W_i\Delta W_j] \\
&= \sum_i \mathbb{E}[e_i^2](t_{i+1} - t_i) \\
&= \mathbb{E}\left[\sum_i e_i^2(t_{i+1} - t_i)\right] \\
&= \mathbb{E}\left[\int_0^T \phi^2(t,\cdot)\mathrm{d}t\right].
\end{aligned}$$

□

Remark 2.3 (alternative method of proof)**.** *Instead of using the independent increments property of BM, one could also insert a conditional expectation (using the tower property thereof) and rely on the martingale property of BM. For details, see Section 4.1, the proof of the isometry property in [Kle07].*

We now extend the definition of the Itô integral from elementary functions to functions in $\mathcal{V}$.

Lemma 2.4. *Let $g \in \mathcal{V}$. Then there exist elementary functions $\phi_n \in \mathcal{V}$ ($n \in \mathbb{N}$) such that*

$$\mathbb{E}\left[\int_0^T (g - \phi_n)^2\mathrm{d}t\right] \overset{n\to\infty}{\longrightarrow} 0.$$

Proof. We will proceed in 3 steps.

Step 1: For $g \in \mathcal{V}$ bounded such that $t \mapsto g(t,\omega)$ is continuous define the elementary function $\phi_n(t,\omega) := \sum_i g(t_i,\omega)\mathbb{1}_{[t_i,t_{i+1})}(t)$. Then

$$\int_0^T (g-\phi_n)^2 \mathrm{d}t \overset{n\to\infty}{\longrightarrow} 0, \quad \forall\omega,$$

since $t \mapsto g(t,\omega)$ is continuous for all $\omega \in \Omega$. By bounded convergence (for this we require g to be bounded), the assertion follows.

Step 2: Suppose that $g \in \mathcal{V}$ (therefore progressively measurable) is bounded. For $t \in [0,T]$, we define

$$F(t) := \int_0^t g(s,\cdot)\mathrm{d}s, \quad G_n(t) := n[F(t) - F(t-\frac{1}{n})^+].$$

F is bounded, hence so is G_n. Furthermore, G_n is continuous and progressively measurable (property is inherited from g). To see that one really has $G_n \in \mathcal{V}$, calculate

$$\begin{aligned}
\mathbb{E}\left[\int_0^T G_n^2(t,\cdot)\mathrm{d}t\right] &= \mathbb{E}\left[\int_0^T \left(n\int_{(t-\frac{1}{n})^+}^t g(s,\cdot)\mathrm{d}s\right)^2 \mathrm{d}t\right] \\
&= n^2\mathbb{E}\left[\int_0^T \left(\int_{(t-\frac{1}{n})^+}^t \underbrace{g(s,\cdot)}_{\le M}\mathrm{d}s\right)^2 \mathrm{d}t\right] \\
&\le n^2 \int_0^T \left(\frac{1}{n}\cdot M\right)^2 \mathrm{d}t \\
&= M^2 T < \infty.
\end{aligned}$$

By virtue of Step 1, there exists, for each $n \in \mathbb{N}$, a sequence of simple processes $(\phi_n^m)_{m\in\mathbb{N}}$ such that

$$\mathbb{E}\left[\int_0^T (G_n - \phi_n^m)^2 \mathrm{d}t\right] \overset{m\to\infty}{\longrightarrow} 0. \tag{2.2}$$

Consider the following sets:

$$\begin{aligned}
\mathcal{A} &:= \left\{(t,\omega) \in [0,T]\times\Omega \,\middle|\, \lim_{n\to\infty} G_n(t,\omega) = g(t,\omega)\right\}^c \in \mathcal{B}([0,T])\otimes\mathcal{F}, \\
\mathcal{A}_\omega &:= \{t \in [0,T] \mid (t,\omega)\in\mathcal{A}\} \in \mathcal{B}([0,T]) \text{ (by Fubini)}.
\end{aligned}$$

Observe that $t \in \mathcal{A}_\omega^c$ whenever $G_n(t,\omega) \to g(t,\omega)$. For all ω we have

$$\lim_{n\to\infty} G_n(t,\omega) = \lim_{n\to\infty} \frac{F(t,\omega) - F((t-\frac{1}{n})^+,\omega)}{\frac{1}{n}} = g(t,\omega), \qquad t > 0,$$

hence the Lebesgue measure of $\mathcal{A}_\omega$ is $\lambda(\mathcal{A}_\omega) = 0$. By Fubini, we therefore have $(\mathbb{P}\otimes\lambda)(\mathcal{A}) = 0$. As g is (by assumption) bounded, we can apply the dominated convergence, which implies that

$$\mathbb{E}\left[\int_0^T (G_n - g)^2 \mathrm{d}t\right] \overset{n\to\infty}{\longrightarrow} 0. \tag{2.3}$$

Combining (2.2) and (2.3), we find that

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\left[\int_0^T (\phi_n^{m_n}(t) - g(t))^2 \mathrm{d}t\right] \\
&\leq 2 \lim_{n\to\infty} \left(\mathbb{E}\left[\int_0^T (G_n(t,\cdot) - g(t,\cdot))^2 \mathrm{d}t\right] + \mathbb{E}\left[\int_0^T (G_n(t,\cdot) - \phi_n^{m_n}(t,\cdot))^2 \mathrm{d}t\right]\right) \\
&= 0,
\end{aligned}$$

where $m_n := \inf\left\{ m \in \mathbb{N} \,\middle|\, \mathbb{E}\left[\int_0^T (G_n(t,\cdot) - \phi_n^m(t,\cdot))^2 \mathrm{d}t\right] < \frac{1}{n}\right\}$.

Step 3: Now assume that $g \in \mathcal{V}$ with no further restrictions – in particular, g need not be bounded. For $n \in \mathbb{N}$ we define

$$h_n(t,\omega) := \begin{cases} -n, & \text{if } g(t,\omega) < -n, \\ g(t,\omega) & \text{if } -n \leq g(t,\omega) \leq n, \\ n & \text{if } g(t,\omega) > n. \end{cases}$$

These functions are bounded and belong to $\mathcal{V}$. Along the lines of Step 2, we define

$$\begin{aligned}
\mathcal{A}^n &:= \{(t,\omega) \in [0,T] \times \Omega \mid \ |g(t,\omega)| > n\}, \\
\mathcal{A}^n_\omega &:= \{t \in [0,T] \mid \ (t,\omega) \in \mathcal{A}^n\}.
\end{aligned}$$

We have $\lim_{n\to\infty} \lambda(\mathcal{A}^n_\omega) = 0$ because g is (by assumption) square-integrable. Therefore,

$$\mathbb{E}\left[\int_0^T (g - h_n)^2 \mathrm{d}t\right] \leq \mathbb{E}\left[\int_{\mathcal{A}^n_\omega} g^2 \mathrm{d}t\right] \to 0$$

as $n \to \infty$. From the first two steps, we have that

$$\mathbb{E}\left[\int_0^T (\phi_n^{m_n} - h_n(t))^2 \mathrm{d}t\right] \leq \frac{1}{n}$$

for a sequence $(m_n)_{n\in\mathbb{N}}$, hence

$$\mathbb{E}\left[\int_0^T (g(t) - \phi_n^{m_n})^2 \mathrm{d}t\right] \to 0,$$

which completes the proof.

□

Corollary 2.5. *For $g \in \mathcal{V}$ let $(\phi_n)_{n\in\mathbb{N}}$ be the sequence from Lemma 2.4. Then the sequence $(\int_0^T \phi_n \mathrm{d}W)_{n\in\mathbb{N}}$ converges in $L^2(\mathbb{P})$.*

Proof. We have seen that $\phi_n \to g$ in $L^2(\mathbb{P} \otimes \lambda)$, implying that it is a Cauchy sequence in this space. From this we infer by Itô's isometry that $(\int_0^T \phi_n \mathrm{d}W)_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\mathbb{P})$:

$$\mathbb{E}\left[\left(\int_0^T \phi_m(t,\cdot)\mathrm{d}W_t - \int_0^T \phi_n(t,\cdot)\mathrm{d}W_t\right)^2\right] = \mathbb{E}\left[\int_0^T (\phi_m(t,\cdot) - \phi_n(t,\cdot))^2 \mathrm{d}t\right] \overset{m,n\to\infty}{\longrightarrow} 0.$$

Since $L^2(\mathbb{P})$ is complete, $(\int_0^T \phi_n \mathrm{d}W)_{n\in\mathbb{N}}$ must converge in that space. □

With the above results, we can define the Itô integral for any $f \in \mathcal{V}$ as

$$I[f] := \int_0^T f(t,\cdot)\mathrm{d}W_t := \lim_{n\to\infty} \int_0^T \phi_n(t,\cdot)\mathrm{d}W_t, \tag{2.4}$$

where the limit is taken in $L^2(\mathbb{P})$ and $\phi_n \to f$ in $L^2(\mathbb{P}\otimes\lambda)$ are elementary functions approximating f.

Remark 2.6. *The above definition is independent of the approximating sequence. To see this, let (ϕ_n) and (ψ_n) be two sequences of elementary functions in $\mathcal{V}$, both approximating $f \in \mathcal{V}$ in L^2-sense. Then by Itô's isometry*

$$\mathbb{E}\left[\left(\int_0^T \phi_n(t,\cdot)\mathrm{d}W_t - \int_0^T \psi_n(t,\cdot)\mathrm{d}W_t\right)^2\right] = \mathbb{E}\left[\int_0^T (\phi_n(t,\cdot)-\psi_n(t,\cdot))^2\mathrm{d}t\right] \overset{n\to\infty}{\longrightarrow} 0,$$

hence both approximating sequences give the same integral in the limit.

The Itô isometry does not only hold for elementary functions, but also for any functions in $\mathcal{V}$:

Corollary 2.7. *Let $f \in \mathcal{V}$. Then*

$$\mathbb{E}\left[\left(\int_0^T f(t,\cdot)\mathrm{d}W_t\right)^2\right] = \mathbb{E}\left[\int_0^T f^2(t,\cdot)\mathrm{d}t\right].$$

Proof. Let (ϕ_n) be a sequence of elementary functions approximating f in L^2-sense. Then

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T f(t,\cdot)\mathrm{d}W_t\right)^2\right] &= \lim_{n\to\infty}\mathbb{E}\left[\left(\int_0^T \phi_n(t,\cdot)\mathrm{d}W_t\right)^2\right] \\ &= \lim_{n\to\infty}\mathbb{E}\left[\int_0^T \phi_n^2(t,\cdot)\mathrm{d}t\right] \\ &= \mathbb{E}\left[\int_0^T f^2(t,\cdot)\mathrm{d}t\right],\end{aligned}$$

where the first and third equality hold by definition of the stochastic integral and the second equality is Itô's isometry for elementary functions (Lemma 2.2). □

An immediate consequence of the isometry is the following result:

Corollary 2.8. *Let $f \in \mathcal{V}$ and $(f_n)_{n\in\mathbb{N}} \subset \mathcal{V}$ such that*

$$\mathbb{E}\left[\int_0^T (f_n(t,\cdot) - f(t,\cdot))^2\,\mathrm{d}t\right] \overset{n\to\infty}{\longrightarrow} 0.$$

Then

$$\int_0^T f_n(t,\cdot)\mathrm{d}W_t \overset{n\to\infty}{\longrightarrow} \int_0^T f(t,\cdot)\mathrm{d}W_t \quad \text{in } L^2(\mathbb{P}).$$

Let's calculate a specific stochastic integral:

Example 2.9. *We want to show that the following formula, which we have already encountered in pathwise Itô calculus, also holds for Itô's stochastic integration:*

$$\int_0^t W_s \mathrm{d}W_s = \frac{1}{2}W_t^2 - \frac{1}{2}t. \tag{2.5}$$

Let $\tau := (t_j)_j$ *be a partition of* $[0,t]$ *and write* $W_j := W_{t_j}$ *for short. We approximate the BM* W *with the following elementary functions:*

$$\phi_n(s,\omega) = \sum_j W_j(\omega)\mathbb{1}_{[t_j,t_{j+1})}(s).$$

With this we have indeed that

$$\begin{aligned}
\mathbb{E}\left[\int_0^t (\phi_n(s,\cdot) - W_s)^2 \, \mathrm{d}s\right] &= \mathbb{E}\left[\sum_j \int_{t_j}^{t_{j+1}} (W_j - W_s)^2 \mathrm{d}s\right] \\
&= \sum_j \int_{t_j}^{t_{j+1}} \mathbb{E}[(W_j - W_s)^2]\mathrm{d}s \\
&= \sum_j \int_{t_j}^{t_{j+1}} (s - t_j) \, \mathrm{d}s \\
&= \sum_j \left[\frac{1}{2}s^2 - t_j s\right]_{s=t_j}^{s=t_{j+1}} \\
&= \sum_j \frac{1}{2}(t_{j+1} - t_j)^2 \ \leq \frac{1}{2}t \cdot \underbrace{\sup_j |t_{j+1} - t_j|}_{=\|\tau\|} \\
&\xrightarrow{\|\tau\|\to 0} 0.
\end{aligned}$$

From Corollary 2.8 we infer (with the shorthand notation $\Delta W_j = W_{j+1} - W_j$*) that*

$$\int_0^t W_s \mathrm{d}W_s = \lim_{\|\tau\|\to 0} \int_0^t \phi_n \mathrm{d}W_s = \lim_{\|\tau\|\to 0} \sum_j W_j \Delta W_j.$$

Observe that

$$\begin{aligned}
\Delta(W_j^2) &= W_{j+1}^2 - W_j^2 \text{ (by definition)} \\
&= (W_{j+1} - W_j)^2 + 2W_j(W_{j+1} - W_j) \\
&= (\Delta W_j)^2 + 2W_j\Delta W_j,
\end{aligned}$$

hence

$$\sum_j (\Delta W_j)^2 + 2\sum_j W_j \Delta W_j = \sum_j \Delta(W_j^2) = W_t^2 - \underbrace{W_0^2}_{=0} = W_t^2.$$

Rearrangement gives

$$\sum_j W_j \Delta W_j = \frac{1}{2}W_t^2 - \frac{1}{2}\sum_j (\Delta W_j)^2.$$

Knowing the quadratic variation of BM is $\langle W\rangle_t = t$, we get the L^2-convergence

$$\sum_j (\Delta W_j)^2 \to t, \quad \text{as } \|\tau\| \to 0,$$

from which (2.5) *follows.*

Exercise 5. *Show that $\int_0^t s \mathrm{d}W_s = tW_t - \int_0^t W_s \mathrm{d}s$.*
Hint: $\sum_j \Delta(s_j W_j) = \sum_j s_j \Delta W_j + \sum_j W_{j+1} \Delta s_j$.

Exercise 6. *Calculate $\mathbb{E}\left[\int_0^t W_s \mathrm{d}W_s\right]$ and $\mathbb{E}\left[\left(\int_0^t W_s \mathrm{d}W_s\right)^2\right]$.*

2.1.2 Properties of the stochastic integral

Theorem 2.10. *Let $f, g \in \mathcal{V}$, $c \in \mathbb{R}$ and $T > 0$.*

(i) For any $0 \leq S \leq T$ we have

$$\int_0^T f \mathrm{d}W = \int_0^S f \mathrm{d}W + \int_S^T f \mathrm{d}W.$$

(ii) The stochastic integral is linear:

$$\int_0^T (cf + g) \mathrm{d}W = c \int_0^T f \mathrm{d}W + \int_0^T g \mathrm{d}W.$$

(iii) $\mathbb{E}[\int_0^T f \mathrm{d}W] = 0$, i.e. the expected value of any Itô integral is zero.

(iv) $\int_0^T f \mathrm{d}W$ is $\mathcal{F}_T$-measurable.

Proof. These properties can be verified for elementary functions. By taking limits, they hold for all functions in $\mathcal{V}$. For (iv) recall that the limit of measurable functions is measurable. □

Theorem 2.11. *Let $f \in \mathcal{V}$. Then the stochastic process $(\int_0^t f \mathrm{d}W)_{t \geq 0}$ is an $(\mathcal{F}_t)$-martingale.*

Proof. We will first prove the assertion for elementary function ϕ_n approximating f and then use the isometry to conclude. Suppressing the dependency on ω in the notation, we let

$$I_n(t) := \int_0^t \phi_n(s) \mathrm{d}W_s.$$

Measurability is clear and the limit of measurable functions is again measurable. For the martingale property let $s > t$. Then

$$\begin{aligned}
\mathbb{E}[I_n(s)|\mathcal{F}_t] &\overset{(i)}{=} \mathbb{E}\left[\int_0^t \phi_n \mathrm{d}W|\mathcal{F}_t\right] + \mathbb{E}\left[\int_t^s \phi_n \mathrm{d}W|\mathcal{F}_t\right] \\
&\overset{(iv)}{=} \int_0^t \phi_n \mathrm{d}W + \mathbb{E}\left[\int_t^s \phi_n \mathrm{d}W|\mathcal{F}_t\right] \\
&= \int_0^t \phi_n \mathrm{d}W + \mathbb{E}\left[\sum e_j^{(n)} \Delta W_j^{(n)}|\mathcal{F}_t\right],
\end{aligned}$$

where the sum is over $t \leq t_j^{(n)} < t_{j+1}^{(n)} \leq s$.
Since $e_j^{(n)}$ is $\mathcal{F}_{t_j^{(n)}}$-measurable and W has independent increments, we have

$$\mathbb{E}\left[e_j^{(n)} \Delta W_j^{(n)}|\mathcal{F}_t\right] = \mathbb{E}\Big[e_j^{(n)} \underbrace{\mathbb{E}[\Delta W_j^{(n)}|\mathcal{F}_{t_j^{(n)}}]}_{=\mathbb{E}[\Delta W_j^{(n)}]=0} |\mathcal{F}_t\Big] = 0,$$

hence we have the martingale property for elementary functions:

$$\mathbb{E}[I_n(s)|\mathcal{F}_t] = \int_0^t \phi_n \mathrm{d}W = I_n(t).$$

Taking the L^2-limit, we get

$$\lim_{n\to\infty} \mathbb{E}[I_n(s)|\mathcal{F}_t] = \lim_{n\to\infty} \int_0^t \phi_n \mathrm{d}W \stackrel{\text{def.}}{=} \int_0^t f \mathrm{d}W.$$

Jensen's inequality allows us to conclude:

$$\begin{aligned}
\mathbb{E}\left[(\mathbb{E}[I_n(s)|\mathcal{F}_t] - \mathbb{E}[I(s)|\mathcal{F}_t])^2\right] &= \mathbb{E}\left[(\mathbb{E}[I_n(s) - I(s)|\mathcal{F}_t])^2\right] \\
&\stackrel{\text{Jensen}}{\leq} \mathbb{E}\left[\mathbb{E}[I_n(s) - I(s)]^2|\mathcal{F}_t\right] \\
&= \mathbb{E}\left[(I_n(s) - I(s))^2\right] \ \text{(tower ppty)} \\
&\xrightarrow{\text{def.}} 0.
\end{aligned}$$

Finally, from the square integrability of $f \in \mathcal{V}$ we infer (by Itô's isometry) that of $I[f]$, hence $I[f]$ is also integrable[2]. □

The following result will be required to show that the Itô integral $\int_0^t f(s,\omega)\mathrm{d}W_s(\omega)$ is (a version of) a continuous function in time t.

Theorem 2.12 (Doob's martingale L^p-inequality)**.** *Let $(M_t)_{t\geq 0}$ be a martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ with almost surely continuous sample paths. Then for all $p \geq 1$, $T \geq 0$ and $\lambda > 0$ we have*

$$\mathbb{P}\left[\sup_{0\leq t\leq T} |M_t| > \lambda\right] \leq \frac{1}{\lambda^p}\mathbb{E}[|M_T|^p].$$

Proof. See Section II §1 in [RY99] or Theorem 4.2 in [SKY01]. The idea is to go from the discrete-time version of the inequality to the continuous-time version via Fatou's lemma (passing to the limit). □

Theorem 2.13. *Let $f \in \mathcal{V}$. Then there exists a modification of the process $(\int_0^t f\mathrm{d}W)_{0\leq t\leq T}$ that is continuous in t.*

Proof. Let (ϕ_n) be a sequence of elementary functions approximating f, i.e.,

$$\mathbb{E}\left[\int_0^T (f - \phi_n)^2 \mathrm{d}t\right] \xrightarrow{n\to\infty} 0.$$

Let

$$I_n(t) := \int_0^t \phi_n(s)\mathrm{d}W_s \quad \text{and} \quad I_t := \int_0^t f(s)\mathrm{d}W_s.$$

We have already seen that $(I_n(t))$ is an $(\mathcal{F}_t)$-martingale and its sample paths are continuous[3] for all $n \in \mathbb{N}$. Consequently, for all $n, m \in \mathbb{N}$, $I_n - I_m$ is also an $(\mathcal{F}_t)$-martingale with continuous sample paths, so Doob's martingale inequality yields

$$\begin{aligned}
\mathbb{P}\left[\sup_{0\leq t\leq T} |I_n(t) - I_m(t)| > \varepsilon\right] &\leq \frac{1}{\varepsilon^2}\mathbb{E}\left[|I_n(T) - I_m(T)|^2\right] \\
&= \frac{1}{\varepsilon^2}\mathbb{E}\left[\int_0^T (\phi_n - \phi_m)^2 \mathrm{d}t\right] \\
&\xrightarrow{m,n\to\infty} 0
\end{aligned}$$

[2]Hölder inequality implies for any random variable X that $\mathbb{E}[|X|] = \mathbb{E}[|X \cdot 1|] \leq \mathbb{E}[1^2]^{\frac{1}{2}} \cdot \mathbb{E}[|X|^2]^{\frac{1}{2}} = \mathbb{E}[|X|^2]^{\frac{1}{2}}$.

[3]This follows from the fact that sample paths of BM are almost surely continuous.

We can therefore choose a subsequence $(n_k) \subset \mathbb{N}$ with $\lim_{k\to\infty} n_k = \infty$ such that

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

The (first) lemma of Borel-Cantelli[4] implies that

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

Consequently, for almost all $\omega \in \Omega$, there exists $k^* = k^*(\omega)$ such that

$$\sup_{0\le t\le T} |I_{n_{k+1}}(t,\omega) - I_{n_k}(t,\omega)| \le 2^{-k} \quad \text{ for all } k \ge k^*$$

Hence for allmost all $\omega \in \Omega$, the sequence $(I_{n_k}(t,\omega))_{k\in\mathbb{N}}$ is uniformly convergent for $t \in [0,T]$ to a limit, which we shall denote by $J_t = J_t(\omega)$. By uniform convergence, the limit is almost surely continuous (in t).
Since $I_{n_k}(t) \to I(t)$ in $L^2(\mathbb{P})$ for all $t \in [0,T]$, we have a.s. convergence along a subsequence, implying that

$$I_t = J_t \quad \mathbb{P}-\text{a.s.} \quad \text{ for all } t \in [0,T].$$

□

2.1.3 Extension of the Itô integral and local martingales

Consider the following class of function:

Definition 2.14. *Let $\mathcal{W}$ be the class of all functions $f\colon [0,\infty) \times \Omega \to \mathbb{R}$ such that*

i) the map $(t,\omega) \to f(t,\omega)$ is progressively measurable;

ii) $\mathbb{P}\left(\int_0^T f^2(t,\omega)\mathrm{d}t < \infty\right) = 1$.

Exercise 7. *Show that $\mathcal{V} \subset \mathcal{W}$.*

For $f \in \mathcal{W}$ we define the stochastic integral as follows:

Step 1: For $n \in \mathbb{N}$ let $\tau_n := \inf\left\{t \ge 0 \,\middle|\, \int_0^t f^2(s,\cdot)\mathrm{d}s \ge n\right\} \wedge T$. This is a stopping time and we have $\lim_{n\to\infty} \mathbb{P}(\tau_n = T) = 1$. Now define $f_n := f\mathbb{1}_{[0,\tau_n]}$. We have

$$\mathbb{E}\left[\int_0^T f_n^2(t,\omega)\mathrm{d}t\right] = \mathbb{E}\left[\int_0^{\tau_n} f^2(t,\omega)\mathrm{d}t\right] < \infty,$$

hence $f_n \in \mathcal{V}$.

[4] If a sequence of events (A_n) satisfies $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.

Step 2: Let $N(\omega) := \inf\{n \in \mathbb{N} \mid \tau_n(\omega) = T\}$. Since

$$\bigcup_{n=1}^{\infty} \{\omega \mid \tau_n = T\} = \left\{\omega \,\middle|\, \int_0^T f^2(t,\omega)\mathrm{d}t < \infty\right\}$$

and the latter has probability 1 by assumption, we have $N < \infty$ $\mathbb{P}$-almost surely. Let $\Omega_0 := \left\{\omega \,\middle|\, t \mapsto \int_0^t f_n \mathrm{d}W \text{ is continuous for all } n\right\}$. By Theorem 2.13, $\mathbb{P}(\Omega_0) = 1$. Let $\Omega_1 := \Omega_0 \cap \{N < \infty\}$ Then we define the Itô integral of f as

$$\int_0^t f(s,\omega)\mathrm{d}W_s(\omega) := \mathbb{1}_{\Omega_1} \int_0^t f_{N(\omega)}(s,\omega)\mathrm{d}W_s(\omega).$$

The integral has almost surely continuous paths and for $n \to \infty$ we have $\int_0^t f_n \mathrm{d}W \to \int_0^t f \mathrm{d}W$ almost surely.

The price we have to pay for the extension from $\mathcal{V}$ to $\mathcal{W}$ is that the integral is not necessarily a martingale any more, but only a so-called *local martingale*. Let us introduce this notion and collect a few useful results. First, let's see why the approach is known as *localization*.

Definition 2.15. *An increasing sequence of stopping times (ν_n) is called a ($\mathcal{V}$-)localizing sequence for f (on $[0,T]$) provided that $f_n(t,\omega) := f(t,\omega)\mathbb{1}_{\{t \le \nu_n\}} \in \mathcal{V}$ for all n and that*

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} \{\omega \mid \nu_n(\omega) = T\}\right) = 1.$$

The sequence (τ_n) from above is a localizing sequence for f.
Now we give two definitions of a local martingale.

Definition 2.16 (Local Martingale 1 – [SKY01] Def. 7.2)**.** *An $(\mathcal{F}_t)$-adapted process (Z_t) is called a* local martingale *w.r.t. $(\mathcal{F}_t)$ if there exists an increasing sequence of $(\mathcal{F}_t)$-stopping times (ν_n) such that $\nu_n \to \infty$ $\mathbb{P}$-almost surely as $n \to \infty$ and*

$$M_t^n := Z_{t\wedge\nu_n} - Z_0$$

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

Definition 2.17 (Local Martingale 2 – [RY99] IV §1, (1.5))**.** *An $(\mathcal{F}_t)$-adapted right-continuous process (Z_t) is called a* local martingale *w.r.t. $(\mathcal{F}_t)$ if there exists an increasing sequence of $(\mathcal{F}_t)$-stopping times (ν_n) such that $\nu_n \to \infty$ $\mathbb{P}$-almost surely as $n \to \infty$ and*

$$M_t^n := Z_{t\wedge\nu_n}\mathbb{1}_{\{\nu_n > 0\}}$$

is a uniformly integrable $(\mathcal{F}_t)$-martingale for every n.

We will work with the first definition.

Remark 2.18. *With those two (different) definitions of local martingales, a few remarks are required:*

- *By choosing $\nu_n \equiv T$ or ∞, we see that every martingale is also a local martingale.*
- *Observe furthermore that integrability of Z is not required. (This implies, that a local martingale is not necessarily a martingale.)*
- *Furthermore, by replacing ν_n by $\nu_n \wedge n$, we go from integrability to uniform integrability, hence the latter holds without loss of generality. (A stopped local martingale is again a local martingale.)*

Corollary 2.19. *For $f \in \mathcal{W}$, the Itô integral $(\int_0^t f \mathrm{d}W)_{t\geq 0}$ is a local martingale.*

Let us now collect the promised results on local martingales.

Lemma 2.20. *A local martingale $Z = (Z_t)_{t\geq 0}$ that is bounded from below (by M) and satisfies $\mathbb{E}[Z_0] < \infty$ is a supermartingale.*

Proof. First observe that if (Z_t) is bounded from below for every $t \geq 0$, then so ist $(Z_t - Z_0)$.
Let (ν_n) be a localizing sequence for Z, i.e. $(Z_{t\wedge\nu_n} - Z_0)$ is a martingale for every n.
The process Z is adapted to the filtration $(\mathcal{F}_t)$ by definition of a local martingale.
For the supermartingale property, we apply Fatou's lemma[5] and get for $s < t$

$$\begin{aligned}
\mathbb{E}\left[Z_t \,\middle|\, \mathcal{F}_s\right] &= \mathbb{E}\left[\liminf_{n\to\infty}(Z_{t\wedge\nu_n} - Z_0) + Z_0 \,\middle|\, \mathcal{F}_s\right] \\
&= \mathbb{E}\left[\liminf_{n\to\infty}(Z_{t\wedge\nu_n} - Z_0) \,\middle|\, \mathcal{F}_s\right] + Z_0 \\
&\overset{\text{Fatou}}{\leq} \liminf_{n\to\infty} \mathbb{E}\left[Z_{t\wedge\nu_n} - Z_0 \,\middle|\, \mathcal{F}_s\right] + Z_0 \\
&= \lim_{n\to\infty} \mathbb{E}\left[Z_{t\wedge\nu_n} - Z_0 \,\middle|\, \mathcal{F}_s\right] + Z_0 \\
&\overset{(*)}{=} Z_s,
\end{aligned}$$

where $(*)$ holds because the stopped process is a martingale.
Finally, integrability holds because (by the tower property) we have

$$\begin{aligned}
Z_0 &\geq \mathbb{E}\left[Z_t \,\middle|\, \mathcal{F}_0\right] \geq \mathbb{E}\left[M \,\middle|\, \mathcal{F}_0\right], \\
\Rightarrow \mathbb{E}\left[Z_0\right] &\geq \mathbb{E}\left[Z_t\right] \geq \mathbb{E}\left[M\right].
\end{aligned}$$

We see that the lower bound may be either a constant or an integrable random variable for the result to hold. ☐

Alternatively one could have assumed $Z_0 = 0$ and $Z_t \geq 0$ before starting with the conditional expectations, but with this assumption, one gets $\mathbb{E}[Z_t] = \mathbb{E}[Z_0]$ for all t, hence we even have a martingale. More generally, we can use dominated convergence to show the following result:

Corollary 2.21. *Let $Z = (Z_t)_{t\geq 0}$ be a continuous local martingale with $|Z_t| \leq M$ for all $t \geq 0$ for some constant M. Then Z is a martingale.*

[5] Fatou's lemma requires that the sequence of random variables be nonnegative. If we merely have $X_n \geq Y$ for an integrable r.v. Y, then we can apply Fatou to $X_n + Y$.

Lemma 2.22. *Let (Z_t) be a local martingale with localizing sequence (ν_n). Then it is a martingale if one of the following conditions holds:*

(i) The process (Z_t) satisfies

$$\sup_{0\leq s\leq t} |Z_s| \in L^1, \quad \forall t > 0.$$

(ii) The stopped sequence $(Z_{t\wedge\nu_n})$ is uniformly integrable, i.e.,

$$\lim_{K\to\infty} \sup_t \mathbb{E}\left[|Z_{t\wedge\nu_n}|\mathbb{1}_{\{|Z_{t\wedge\nu_n}|\geq K\}}\right] = 0$$

Proof. (i) From the local martingale property we have

$$\mathbb{E}\left[Z_{t\wedge\nu_n} \mid \mathcal{F}_s\right] = Z_{s\wedge\nu_n}. \tag{2.6}$$

Property (i) gives

$$|Z_{t\wedge\nu_n}| \leq \sup_{0\leq s\leq t} |Z_s| \in L^1, \quad \forall t > 0.$$

Hence, by Lebesgue's dominated convergence theorem we get

$$\mathbb{E}\left[Z_t \mid \mathcal{F}_s\right] = \mathbb{E}\left[\lim_{n\to\infty} Z_{t\wedge\nu_n} \,\middle|\, \mathcal{F}_s\right] = \lim_{n\to\infty} \mathbb{E}\left[Z_{t\wedge\nu_n} \mid \mathcal{F}_s\right] = \lim_{n\to\infty} Z_{s\wedge\nu_n} = Z_s,$$

hence we have the martingale property. Adaptedness and integrability holds by assumption.

(ii) We have $Z_{t\wedge\nu_n} \stackrel{n\to\infty}{\longrightarrow} Z_t$ $\mathbb{P}$-a.s. and uniform integrability of $(Z_{t\wedge\nu_n})$, hence we have (conditional) L^1-convergence:

$$\lim_{n\to\infty} \mathbb{E}\left[Z_{t\wedge\nu_n} \mid \mathcal{F}_s\right] = \mathbb{E}\left[Z_t \mid \mathcal{F}_s\right]$$

and the result follows as before.

□

2.2 Itô Processes

As before, assume that W is a 1-dimensional standard BM on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

Definition 2.23. *A stochastic process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ of the form*

$$X_t = X_0 + \int_0^t a_s \mathrm{d}s + \int_0^t b_s \mathrm{d}W_s \tag{2.7}$$

where $b \in \mathcal{W}$ and a is adapted and satisfies the integrability condition

$$\mathbb{P}\left(\int_0^t |a_s|\mathrm{d}s < \infty \ \forall t \geq 0\right) = 1$$

is called an Itô process.

Equation (2.7) written in differential form reads

$$\mathrm{d}X_t = a_t \mathrm{d}t + b_t \mathrm{d}W_t.$$

With this shorthand notation, equation (2.5) can be written as

$$\mathrm{d}(\frac{1}{2}W_t^2) = W_t \mathrm{d}W_t + \frac{1}{2}\mathrm{d}t.$$

Example 2.24. *From Financial Mathematics I we know already the simplest possible Itô process, namely a BM with drift $\mu \in \mathbb{R}$ and volatility $\sigma > 0$:*

$$\mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t.$$

If X is an Itô process, then how do we define the integral $\int_0^T f(t,\omega)\mathrm{d}X_t(\omega)$?

Definition 2.25. *Let $f\colon [0,\infty) \times \Omega \to \mathbb{R}$ be progressively measurable and let X be an Itô process with drift a and volatility b as stated above. Assume that*

$$\int_0^T |f_s a_s|\mathrm{d}s < \infty, \quad \int_0^T |f_s b_s|^2 \mathrm{d}s < \infty \quad \mathbb{P} - \textit{a.s.}$$

Then define for $t \in [0,T]$:

$$\int_0^t f_s \mathrm{d}X_s := \int_0^t f_s a_s \mathrm{d}s + \int_0^t f_s b_s \mathrm{d}W_s.$$

Before we look at the famous Itô formula, let us see what the quadratic variation of an Itô process is.

Lemma 2.26. *Let X be an Itô process with drift a and volatility b. Then $\langle X\rangle_t = \int_0^t (b_s)^2 \mathrm{d}s$ $\mathbb{P}$-a.s. $\forall t \in [0,T]$.*

Exercise 8. *Prove Lemma 2.26. Follow the steps in [SKY01] (Section 8.6, Theorem 8.6).*

It is easy to remember this Lemma's statement in its differential form by following these rules:

$$\mathrm{d}t \cdot \mathrm{d}t = \mathrm{d}t \cdot \mathrm{d}W_t = \mathrm{d}W_t \cdot \mathrm{d}t = 0, \quad \mathrm{d}W_t \cdot \mathrm{d}W_t = \mathrm{d}t.$$

Applying them to the Itô process X we get

$$\begin{aligned} \mathrm{d}X_t \cdot \mathrm{d}X_t &= (a_t \mathrm{d}t + b_t \mathrm{d}W_t) \cdot (a_t \mathrm{d}t + b_t \mathrm{d}W_t) \\ &= (a_t)^2 (\mathrm{d}t)^2 + 2 a_t b_t \mathrm{d}t \mathrm{d}W_t + (b_t)^2 (\mathrm{d}W_t)^2 \\ &= (b_t)^2 \mathrm{d}t. \end{aligned}$$

Theorem 2.27 (Itô formula)**.** *Let X be an Itô process and let $Y = f(t, X)$ for a function f that is continuously differntiable w.r.t. the time variable and twice continuously differentiable w.r.t. the space variable. Then Y is an Itô process with*

$$\mathrm{d}Y_t = \frac{\partial}{\partial t} f(t, X_t)\mathrm{d}t + \frac{\partial}{\partial x} f(t, X_t)\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2}{\partial x^2} f(t, X_t)\mathrm{d}\langle X\rangle_t. \tag{2.8}$$

Example 2.28 (Geometric Brownian motion)**.** *For $\mu \in \mathbb{R}$ and $\sigma > 0$ and a standard BM W let*

$$S_t = S_0 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right).$$

We can write this process (known from the Black-Scholes option pricing model) as an Itô process: Let $S_t = S_0 \cdot F(t, W_t)$ with $F(t,x) = \exp\left(\sigma x + \left(\mu - \frac{1}{2}\sigma^2\right) t\right)$. The above Itô formula tells us that

$$\begin{aligned} \mathrm{d}S_t &= \left(\mu - \frac{1}{2}\sigma^2\right) S_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t + \frac{1}{2}\sigma^2 S_t \mathrm{d}t \\ &= S_t(\mu \mathrm{d}t + \sigma \mathrm{d}W_t) \end{aligned}$$

Example 2.29. *We have already seen in an exercise that*

$$\int_0^t s \mathrm{d}W_s = tW_t - \int_0^t W_s \mathrm{d}s. \tag{2.9}$$

We can also show this by applying Itô's formula to $Y_t = f(t, W_t)$ with $f(t,x) = tx$. The formula gives

$$d(tW_t) = dY_t = W_t \mathrm{d}t + t\mathrm{d}W_t,$$

hence (with $Y_0 = 0$)

$$tW_t = \int_0^t W_s \mathrm{d}s + \int_0^t s \mathrm{d}W_s.$$

Rearrangement gives (2.9).

Example 2.30. *We can use Itô's formula to show that $Y_t = e^{\frac{1}{2}t}\sin(W_t)$ $(t \geq 0)$ is a martingale. To this end, we write Y_t as a stochastic integral: By Itô's formula,*

$$\begin{aligned} d(e^{\frac{1}{2}t}\sin(W_t)) &= \frac{1}{2} \cdot e^{\frac{1}{2}t} \cdot \sin(W_t)\mathrm{d}t + e^{\frac{1}{2}t}\cos(W_t)\mathrm{d}W_t + \frac{1}{2} \cdot e^{\frac{1}{2}t} \cdot (-\sin(W_t))\mathrm{d}t \\ &= e^{\frac{1}{2}t}\cos(W_t)\mathrm{d}W_t, \end{aligned}$$

hence $Y_t = \int_0^t e^{\frac{1}{2}t}\cos(W_t)\mathrm{d}W_t$. By Theorem 2.11, this is an integral if the integrand belongs to $\mathcal{V}$. (This is left as an exercise.)

Remark 2.31. *Itô's formula is proven in the literature by different means:*

- *by applying Taylor's formula (e.g. in [Øk03], proof starting on page 46)*
- *by proving it for polynomial functions, by stopping the process and thus working on compact sets and by then using convergence results (e.g. in [RY99], Chapter IV, §3, (3.3))*

We can also define a multi-dimensional Itô process as follows: We first let $W = (W^1, \dots, W^m)$ denote an m-dimensional BM.
Let each component of $A = (a^i)_{i=1,\dots,n}$ and $B = (b^{ij})_{i=1,\dots,n; j=1,\dots,m}$ satisfy the conditions we previously imposed on a and b, respectively. Then an n-dimensional process $X = (X^1, \dots, X^n)$ is an Itô process if we have

$$dX_t = A\mathrm{d}t + B\mathrm{d}W_t,$$

in other words we have

$$\begin{aligned}
\mathrm{d}X_t^1 &= a_t^1 \mathrm{d}t + b_t^{11} \mathrm{d}W_t^1 + \ldots + b_t^{1m} \mathrm{d}W_t^m, \\
&\vdots \\
\mathrm{d}X_t^n &= a_t^n \mathrm{d}t + b_t^{n1} \mathrm{d}W_t^1 + \ldots + b_t^{nm} \mathrm{d}W_t^m.
\end{aligned}$$

Theorem 2.32. *Let X be an Itô process and let $f(t,x) = (f_1(t,x), \dots, f_r(t,x))$ be a C^1-function in time and a C^2-function in space. Then $Y_t(\omega) := f(t, X_t(\omega))$ defines an Itô process with representation*

$$\mathrm{d}Y_t^k = \frac{\partial}{\partial t} f_k(t, X_t)\mathrm{d}t + \sum_i \frac{\partial}{\partial x_i} f_k(t, X_t)\mathrm{d}X_t^i + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} f_k(t, X_t)\mathrm{d}X_t^i \mathrm{d}X_t^j.$$

for $k \in \{1, \dots, r\}$ and the sums running over $i, j \in \{1, \dots, n\}$.

Theorem 2.33 (Integration by parts)**.** *Let X and Y be Itô processes on $\mathbb{R}$. Then we have*

$$\int_0^t X_s \mathrm{d}Y_s = X_t Y_t - X_0 Y_0 - \int_0^t Y_s \mathrm{d}X_s - \int_0^t \mathrm{d}X_s \mathrm{d}Y_s.$$

The last term is also written as quadratic covariation of X and Y.

Remark 2.34. *The covariation of X and Y can be defined as*

$$\langle X, Y \rangle_t := \lim_{n \to \infty} \sum_{t_i \in \tau_n \cup \{t\}, t_i \le t} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})$$

It exists if and only if $\langle X + Y \rangle$, $\langle X \rangle$ and $\langle Y \rangle$ exist and we can alternatively define it via the polarization formula

$$\langle X, Y \rangle = \frac{1}{2}\left(\langle X + Y \rangle - \langle X \rangle - \langle Y \rangle\right).$$

2.3 The Martingale Representation Theorem

We have already seen in Theorem 2.11 that the stochastic integral of a function $f \in \mathcal{V}$ is a martingale. In this chapter we want to prove that the reverse holds, i.e., that every martingale (minus its expected value) has a (unique) representation as a stochastic integral of some function $f \in \mathcal{V}$.
Let us first state a slightly more general result:

Theorem 2.35 (Itô representation theorem)**.** *Assume that we have a filtered probability space and that its filtration is generated by a standard BM, i.e., $\mathcal{F}_t = \sigma(W_s, s \leq t) \vee \mathcal{N}$. Let $F \in L^2(\mathcal{F}_T)$. Then there exists a unique $f \in \mathcal{V}$ such that*

$$F(\omega) = \mathbb{E}[F] + \int_0^T f(t,\omega)\mathrm{d}W_t(\omega).$$

Before we prove this, let us formulate the martingale representation theorem, which follows directly from the above result.

Theorem 2.36 (Martingale representation theorem (MRT))**.** *Let $M = (M_t)_{t\geq 0}$ be an $(\mathcal{F}_t)$-martingale with $M_t \in L^2(\mathbb{P})$. Then there exists a unique (in $L^2(\mathbb{P}\otimes\lambda)$) process $g \in \mathcal{V}$ such that*

$$M_t(\omega) = \mathbb{E}[M_0] + \int_0^t g(s,\omega)\mathrm{d}W_s(\omega), \quad \mathbb{P}\text{-a.s. for all } t \geq 0.$$

Proof. By applying the Itô representation theorem to $T = t$ and $F = M_t$ and by using that $\mathbb{E}[M_t] = \mathbb{E}[M_0]$ for all $t \geq 0$, we see that there exists $f^{(t)} \in \mathcal{V}$ such that

$$M_t = \mathbb{E}[M_t] + \int_0^t f^{(t)}\mathrm{d}W_s = \mathbb{E}[M_0] + \int_0^t f^{(t)}\mathrm{d}W_s. \tag{2.10}$$

As Itô integrals are martingales, we have that for every $0 \leq t_1 \leq t_2$ the following holds:

$$M_{t_1} = \mathbb{E}\left[M_{t_2} \mid \mathcal{F}_{t_1}\right] \overset{(*)}{=} \mathbb{E}[M_0] + \mathbb{E}\left[\int_0^{t_2} f^{(t_2)}\mathrm{d}W_s \,\middle|\, \mathcal{F}_{t_1}\right] = \mathbb{E}[M_0] + \int_0^{t_1} f^{(t_2)}\mathrm{d}W_s, \tag{2.11}$$

where we used (2.10) for t_2 in equality $(*)$. Equation (2.10) also holds for t_1 instead of t, i.e. we have

$$M_{t_1} = \mathbb{E}[M_0] + \int_0^{t_1} f_s^{(t_1)}\mathrm{d}W_s. \tag{2.12}$$

Comparing (2.11) and (2.12) we get

$$0 = \int_0^{t_1} \left(f_s^{(t_1)} - f_s^{(t_2)}\right)\mathrm{d}W_s.$$

Applying the Itô isometry (and Tonelli's theorem for interchanging integral and expected value) we get

$$0 = \mathbb{E}\left[\left(\int_0^{t_1} \left(f_s^{(t_1)} - f_s^{(t_2)}\right)\mathrm{d}W_s\right)^2\right] = \int_0^{t_1} \mathbb{E}\left[\left(f_s^{(t_1)} - f_s^{(t_2)}\right)^2\right]\mathrm{d}s,$$

hence $f^{(t_1)} = f^{(t_2)}$ for almost all $(t,\omega) \in [0,t_1]\times\Omega$. As $0 \le t_1 \le t_2$ were arbitrarily chosen points, we define $f\colon [0,\infty)\times\Omega$ as

$$f(s,\omega) := f^{(t)}(s,\omega) \quad \text{if } s \in [0,t].$$

Then indeed we have

$$M_t = \mathbb{E}[M_0] + \int_0^t f^{(t)} \mathrm{d}W_s = \mathbb{E}[M_0] + \int_0^t f \mathrm{d}W_s, \quad \text{for all } t \ge 0.$$

□

Example 2.37 (Stochastic Exponential)**.** *For some $h \in L^2[0,T]$ define*

$$Y_t := \mathcal{E}(\int_0^t h_s \mathrm{d}W_s) := \exp\left(\int_0^t h_s \mathrm{d}W_s - \frac{1}{2}\int_0^t (h_s)^2 \mathrm{d}s\right), \quad 0 \le t \le T. \tag{2.13}$$

We have seen in an exercise (sheet 2) that

$$Y_t = 1 + \int_0^t Y_s h_s \mathrm{d}W_s, \quad 0 \le t \le T. \tag{2.14}$$

This is precisely the representation of Y that is suggested in the Itô representation theorem, provided $Yh \in \mathcal{V}$.

Remark 2.38.

1. *We know that if the integrand in the Itô representation belongs to $\mathcal{V}$, then the process is a martingale. In the special case of the stochastic exponential there are a number of results giving sufficient conditions on h such that $\mathcal{E}(\int_0^t h_s \mathrm{d}W_s)$ is a martingale – e.g. Novikov's condition, which requires that $\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T h_s^2 \mathrm{d}s\right)\right] < \infty$. This condition is satisfied for any deterministic function $h \in L^2[0,T]$.*

2. *Consider the stochastic exponential for $h_t \equiv \sigma$. In this case Novikov's condition is clearly satisfied, one can check that $\sigma Y \in \mathcal{V}$.*

   If we multiply (2.14) *by* $\exp\left(\frac{\sigma^2 t}{2}\right)$ *and replace Y by expression* (2.13)*, we get*

   $$\exp(\sigma W_t) = \exp\left(\frac{\sigma^2 t}{2}\right) + \int_0^t \sigma \exp\left(\frac{-\sigma^2(s-t)}{2} + \sigma W_s\right) \mathrm{d}W_s$$

   This implies that the random variable

   $$\widetilde{Y} = \sum_{k=1}^n a_k \exp(\sigma_k W_{t_k})$$

   with arbitrary a_k, σ_k and t_k can be written as an Itô integral with respect to a function in $\mathcal{V}$.

Inspired by the above results we define the following set:

Definition 2.39. *Let $\mathcal{S}$ be the linear span of all random variables that can be written as*

$$\exp\left(\int_0^T h_s \mathrm{d}W_s - \frac{1}{2}\int_0^T (h_s)^2 \mathrm{d}s\right) \quad \text{for } h \in L^2[0,T]. \tag{2.15}$$

We will prove the Itô representation theorem in several steps:

Step 1: Every random variable $Y \in \mathcal{S}$ has an Itô representation.

We have already seen (in Equation (2.14)) that every $Y \in \mathcal{S}$ has the representation

$$Y_t = 1 + \int_0^t Y_s h_s \mathrm{d}W_s, \quad 0 \le t \le T.$$

What is left to check is that for $h \in L^2[0,T]$, the process $(t,\omega) \mapsto Y_t(\omega)h_t$ belongs to $\mathcal{V}$.

Step 2: If a sequence of random variables converging in L^2 has an Itô representation, then the limit has an Itô representation as well.

Proposition 2.40. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables with Itô representations*

$$X_n = \mathbb{E}[X_n] + \int_0^T \phi_t^{(n)} \mathrm{d}W_t, \quad \phi^{(n)} \in \mathcal{V}. \tag{2.16}$$

If $X_n \overset{n\to\infty}{\longrightarrow} X$ in $L^2(\mathbb{P})$, then there exists $\phi \in \mathcal{V}$ such that $\phi^{(n)} \overset{n\to\infty}{\longrightarrow} \phi$ in $L^2(\mathbb{P}\otimes\lambda)$ satisfying

$$X = \mathbb{E}[X] + \int_0^T \phi_t \mathrm{d}W_t. \tag{2.17}$$

Proof. From $L^2(\mathbb{P})$-convergence follows the convergence of the expected values, i.e., $\mathbb{E}[X_n] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X]$. As $(X_n - \mathbb{E}[X_n])_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{L}^2(\mathbb{P})$, we can apply Itô's isometry to see that $(\phi^{(n)})_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\mathbb{P}\otimes\lambda)$:

$$\begin{aligned}
\mathbb{E}[(X_n - X_m)^2] &= \mathbb{E}\left[\left(\mathbb{E}[X_n - X_m] + \int_0^T (\phi_t^{(n)} - \phi_t^{(m)})\mathrm{d}W_t\right)^2\right] \\
&\overset{\text{isom.}}{=} (\mathbb{E}\left[X_n - X_m\right])^2 + \mathbb{E}\left[\int_0^T (\phi_t^{(n)} - \phi_t^{(m)})^2 \mathrm{d}t\right] \\
&= (\mathbb{E}\left[X_n - X_m\right])^2 + \int_0^T \mathbb{E}\left[(\phi_t^{(n)} - \phi_t^{(m)})^2\right]\mathrm{d}t \\
&\overset{m,n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

As the functions $\phi^{(n)}$ belong to $\mathcal{V}$ for each n, which is complete[6], there exists $\phi \in \mathcal{V}$ such that $\phi = \lim_{n\to\infty} \phi^{(n)}$.

Finally, we take $L^2(\mathbb{P})$-limits on both sides of (2.16). Again by Itô's isometry and by using that $\mathbb{E}[X_n] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X]$ we get that

$$X = \lim_{n\to\infty} X_n = \lim_{n\to\infty}\left(\mathbb{E}[X_n] + \int_0^T \phi_t^{(n)}\mathrm{d}W_t\right) = \mathbb{E}[X] + \int_0^T \phi_t \mathrm{d}W_t.$$

□

[6]It is left as an exercise to verify that $\mathcal{V}$ is a closed linear subset of $L^2(\mathbb{P}\otimes\lambda)$ and therefore complete. Let it suffice to remind the reader that if we have a Cauchy sequence in L^2, then the sequence does not only converge in L^2, but there also exists a subsequence which converges almost surely.

Step 3: $\mathcal{S}$ is dense in $L^2(\mathcal{F}_T, \mathbb{P})$.

We need a few results before tackling the main one.
The following theorem follows from the π-λ-theorem. Its proof is left as an exercise; it can be found in [SKY01].

Theorem 2.41 (Monotone Class Theorem)**.** *Let $\mathcal{A}$ be a π-system containing Ω, i.e., for any two sets $A, B \in \mathcal{A}$, the intersection $A \cap B$ belongs to $\mathcal{A}$ as well. Let $\mathcal{H}$ be a collection of functions from Ω to $\mathbb{R}$ with the following properties:*

(i) $\mathcal{H}$ is a vector space;

(ii) if $A \in \mathcal{A}$, then $\mathbb{1}_A \in \mathcal{H}$;

(iii) if $f_n \in \mathcal{H}$ for $n \in \mathbb{N}$ such that $f_n \geq 0$ and $f_n \uparrow f$ for some bounded function f, then $f \in \mathcal{H}$;

Then $\mathcal{H}$ contains all bounded functions that are $\sigma(\mathcal{A})$-measurable.

In the next result we state that such random variables that depend on finitely many times of a standard BM are dense in $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$:

Lemma 2.42 (Lemma 12.3 in [SKY01])**.** *Let $\mathcal{D}$ denote the set of random variables that can be written as*

$$f(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}) \quad \textit{or} \quad f(W_{t_1}, W_{t_2}, \ldots, W_{t_n})$$

for $n \in \mathbb{N}$ and some points $0 = t_0 < t_1 < \ldots < t_n = T$ and a bounded Borel-mb. function $f\colon \mathbb{R}^n \to \mathbb{R}$. Then $\mathcal{D}$ is dense in $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$.

Before we give the proof, let us remark that the first representation has the advantage that all arguments are independent of each other. However, any function with one representation can also be presented in the other way.

Proof. Let $\mathcal{A}$ denote all subsets of Ω of the form

$$A = \{W_{t_1} < x_1, W_{t_2} < x_2, \ldots, W_{t_n} < x_n\} \subset \Omega.$$

Then $\mathcal{A}$ is a π-system and $\Omega \in \mathcal{A}$.
Let $\mathcal{H}_0$ denote the set of all bounded random variables that can be written as limit of a monotone increasing sequence of elements in $\mathcal{D}$. Let $\mathcal{H}$ be the vector space generated by $\mathcal{H}_0$. Then $\mathcal{H}$ has all properties required to apply the monotone class theorem. Furthermore, for any $A \in \mathcal{A}$, we have $\mathbb{1}_A \in \mathcal{H}$. The monotone class theorem tells us that $\mathcal{H}$ contains all bounded $\sigma(\mathcal{A})$-measurable functions. By definition of $\mathcal{F}_T$ (generated by $(W_t)_{t \leq T}$, augmented by the $\mathbb{P}$-nullsets), $\mathcal{F}_T = \sigma(\mathcal{A}) \vee \mathcal{N}$. Hence, for any bounded elementary $\mathcal{F}_T$-measurable X there exists a $\sigma(\mathcal{A})$-measurable Y such that $\mathbb{P}(X \neq Y) = 0$.

$\mathcal{H}$ contains all bounded functions in $L^2(\Omega, \sigma(\mathcal{A}), \mathbb{P})$ and due to the approximation property, it is even a dense subset thereof. Thus, $\mathcal{D}$ is dense in $L^2(\Omega, \mathcal{F}_T, \mathbb{P})$. □

Finally, we recall one more useful result from the theory of Hilbert spaces:

Lemma 2.43. *Set $\mathcal{D}$ be a closed linear subspace of $L^2(\mathbb{P})$ and $\mathcal{S} \subset \mathcal{D}$. If $\mathcal{D} \cap \mathcal{S}^\perp = \{0\}$, then $\overline{\mathcal{S}} = \mathcal{D}$.*

Proof. See Lemma 12.4 in [SKY01]. $\square$

Proposition 2.44. *$\mathcal{S}$ is dense in $L^2(\mathcal{F}_T, \mathbb{P})$.*

Proof of Proposition 2.44. The goal is to show that $\widetilde{\mathcal{S}}$, the set $\mathcal{S}$ restricted to L^2-step functions h, is dense in $\mathcal{D}$. As the latter is dense in $L^2(\mathcal{F}_T, \mathbb{P})$, this completes the proof. Let $g \in \mathcal{D}$ be orthogonal to all functions from $\widetilde{\mathcal{S}}$. In particular, for any $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and $t_1, \ldots, t_n \in [0, T]$,

$$G(a) := \int_\Omega \exp\left(a_1 W_{t_1} + \ldots a_n W_{t_n}\right) g \mathrm{d}\mathbb{P} = 0,$$

where we let $h_k = a_k \mathbb{1}_{[0,t_k]} \in L^2[0,T]$ for $k = 1, \ldots, n$ and the corresponding stochastic exponential

$$\mathcal{E}(\int_0^{t_k} h_k \mathrm{d}W) = \mathcal{E}(a_k W_{t_k}) = \exp\left(a_k W_{t_k} - \frac{a_k^2 t_k}{2}\right) = \exp(a_k W_{t_k}) \cdot c_k,$$

where c_k is a deterministic factor.
As G is real analytic, it has a complex extension to $\mathbb{C}^n$ (by the principle of permanence), i.e.

$$G(z) := \int_\Omega \exp\left(z_1 W_{t_1} + \ldots z_n W_{t_n}\right) g \mathrm{d}\mathbb{P} \equiv 0, \quad z \in \mathbb{C}^n.$$

This holds in particular for $z = ia$ for some $a \in \mathbb{R}^n$. Hence we have for any $\phi \in C_0^\infty(\mathbb{R}^n)$ (dense subset of $\mathcal{D}$):

$$\begin{aligned}
&\int_\Omega \phi(W_{t_1}, \ldots, W_{t_n}) g \mathrm{d}\mathbb{P} \\
&= \int_\Omega (2\pi)^{-n/2} \left\{ \int_{\mathbb{R}^n} \widehat{\phi}(a) \exp\left(i(a_1 W_{t_1} + \ldots + a_n W_{t_n})\right) \mathrm{d}a \right\} g \mathrm{d}\mathbb{P} \\
&= (2\pi)^{-n/2} \int_{\mathbb{R}^n} \widehat{\phi}(a) \left(\int_\Omega e^{i(a_1 W_{t_1} + \ldots + a_n W_{t_n})} g \mathrm{d}\mathbb{P} \right) \mathrm{d}a \\
&= (2\pi)^{-n/2} \int_{\mathbb{R}^n} \widehat{\phi}(a) G(ia) \mathrm{d}a = 0,
\end{aligned}$$

where

$$\widehat{\phi}(a) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \phi(x) e^{-ix \cdot a} \mathrm{d}x$$

and

$$\phi(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \widehat{\phi}(a) e^{ix \cdot a} \mathrm{d}a$$

are the Fourier transform and its inverse. Hence, $g \in \mathcal{D}$ is orthogonal to all functions in C_0^∞, which is a dense subset of $\mathcal{D} \subset L^2$. This can only be true for $g = 0$. Thus, $\mathcal{D} \cap \widetilde{\mathcal{S}}^\perp = \{0\}$. We can directly apply the previous lemma to see that $\widetilde{\mathcal{S}}$ is dense in $\mathcal{D}$, which completes the proof, because if $\widetilde{\mathcal{S}}$ is dense in L^2, then this is certainly true for the larger space $\mathcal{S}$. $\square$

Step 4: The Itô representation is unique.

Let $f, g \in \mathcal{V}$ such that

$$F = \mathbb{E}\left[F\right] + \int_0^T f_t \mathrm{d}W_t \quad \text{and} \quad F = \mathbb{E}\left[F\right] + \int_0^T g_t \mathrm{d}W_t.$$

Then $\int_0^T (f_t - g_t)\mathrm{d}W_t = 0$, which implies that $\mathbb{E}\left[\left(\int_0^T (f_t - g_t)\mathrm{d}W_t\right)^2\right] = 0$. By Itô's isometry, we therefore have

$$\mathbb{E}\left[\int_0^T (f_t - g_t)^2 \mathrm{d}t\right] = 0,$$

which is want we wanted to show.
This completes the proof of Theorem 2.35.

Example 2.45. *Let $F = W_1^2$ and consider the martingale $M_t := \mathbb{E}\left[F \mid \mathcal{F}_t\right]$ with the usual filtration (generated by W, augmented by the nullsets). We have already seen (sheet 2) that the conditional distribution of W_1 given W_t ($t \leq 1$) is $\mathcal{N}(W_t, 1-t)$, hence*

$$M_t = \mathbb{E}\left[W_1^2 \mid \mathcal{F}_t\right] = W_t^2 + (1-t).$$

Itô's formula (for $F(t,x) = x^2$) tells us that

$$W_t^2 = 0 + \int_0^t 2W_s \mathrm{d}W_s + \frac{1}{2}\int_0^t 2\mathrm{d}s = t + 2\int_0^t W_s \mathrm{d}W_s.$$

Therefore, M_t has the representation

$$M_t = 1 + \int_0^t 2W_s \mathrm{d}W_s.$$

Exercise 9. *Find the integral representations of the following random variables:*

a) $F = \int_0^T W_s \mathrm{d}s$,

b) $F = \exp(W_T)$, *(*Hint: What is $\mathrm{d}(\exp(W_t - \frac{1}{2}t))$?*)*

c) $F = \sin(W_T)$. *(*Hint: What is $\mathrm{d}(\exp(\frac{1}{2}t)\sin(W_t))$?*)*

2.3.1 Explicit formula for the integrand in the MRT - An excursion to functional Itô and Malliavin calculus

The MRT only states that an integrand $f \in \mathcal{V}$ exists such that an L^2-martingale M has the representation $M_t = \mathbb{E}[M_0] + \int_0^t f_s \mathrm{d}W_s$. If we know how to approximate M_t by linear combinations of stochastic exponentials, then we even know how to construct the integrand.
In applications, Itô's formula is often useful to find the integrand. As long as we consider stochastic integrals w.r.t. a standard BM W, if $M_t = F(t, W_t)$, then Itô's formula tells us that the integrand should be $\partial_x F(t, W_t)$. Intuitively speaking, if we represent a random variable as an integral, the integrand should be something like the derivative of the random variable.
In his paper *Functional Itô calculus* [Dup09], Bruno Dupire introduced vertical and horizontal derivatives of a process $F(t, X_t)$. This approach is then extended by Rama Cont & David-Antoine Fournié in *Functional Itô calculus and stochastic integral representation of martingales* [CF13] to present explicitely the integrand of the MRT.
We shall only briefly sketch the relevant objects to convey an idea of the topic.
Introduce the notations for horizontal extension and vertical perturbation of a process $x_t = (x(u), 0 \le u \le t)$, which we illustrate in Figure 1:

$$x_{t,h}(u) = x(u), \quad u \in [0,t]; \qquad x_{t,h}(u) = x(t), \quad u \in (t, t+h];$$
$$x_t^h(u) = x(u), \quad u \in [0,t); \qquad x_t^h(t) = x(t) + h.$$

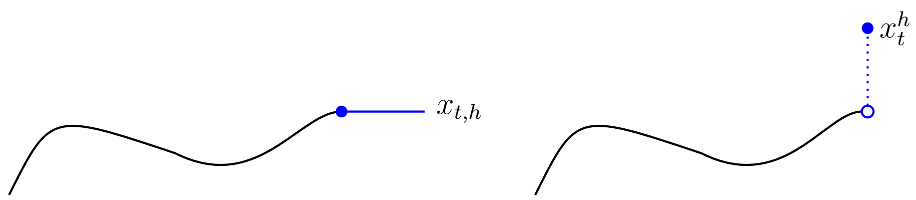

Figure 1: Horizontal extension $x_{t,h}$ and vertical perturbation x_t^h of a path x

With this the *horizontal derivative* of an $\mathbb{R}^d$-valued process F is defined as

$$\mathcal{D}_t F(x,v) := \lim_{h \to 0^+} \frac{1}{h} \left(F_{t+h}(x_{t,h}, v_{t,h}) - F_t(x_t, h_t) \right)$$

and the *vertical derivative* is defined as

$$\nabla_x F(x,v) := (\partial_i F_t(x,v), i = 1, \dots, d) \text{ where } \partial_i F_t(x,v) := \lim_{h \to 0} \frac{1}{h} \left(F_t(x_t^{he^i}, v) - F_t(x,v) \right),$$

if the limits exist and where $(e_i, i = 1, \dots, d)$ is the canonical basis in $\mathbb{R}^d$. If Y has the representation $Y(t) = F_t(X_t, A_t)$, where A is the local quadratic variation[7] of X, then $\nabla_X Y(t) := \nabla_x F_t(X_t, A_t)$ is called the vertical derivative of Y w.r.t. X.
Besides proving a functional Itô formula with the above derivatives (Theorem 4.1), Cont and Fournié prove the following MRT:

[7] $\langle X \rangle_t = \int_0^t A(s) \mathrm{d}s$

Theorem 2.46 (Theorem 5.9 in [CF13]). *For any square-integrable $(\mathcal{F}_t^X)_{t\in[0,T]}$-martingale Y, we have the representation*

$$Y(T) = Y(0) + \int_0^T \nabla_X Y \mathrm{d}X.$$

The (non-pathwise) calculus of derivatives of random variables is known as *Malliavin calculus*. It is concerned with the regularity random variables and gives a notion of the derivative of a random variable or stochastic process – the Malliavin derivative. For further information, consult e.g. [Imk08] or [Bel06].

2.3.2 Application to the Black Scholes Model

In this short section, we will see how the stochastic integral that we have introduced and worked with appears quite naturally in the Black Scholes model and which topics will need to be addressed next in order to have the same machinery available as we had for discrete time financial markets.

Let the price process of a riskless asset be given by $(S_t^0)_{t\geq 0}$, where $S_0^0 = 1$ and $\mathrm{d}S_t^0 = rS_t^0\mathrm{d}t$. In other words, $S_t^0 = e^{rt}$ for any $t \geq 0$. Furthermore, let the price process of a risky asset S be a geometric BM (cf. Example 2.28), i.e., $\mathrm{d}S_t = S_t(\mu\mathrm{d}t + \sigma\mathrm{d}W_t)$. A strategy $\phi = (\phi_t)_{0\leq t\leq T}$ for trading in riskless and risky asset until maturity T is a 2-dimensional process, i.e., $\phi = (H^0, H)$, such that the value of the portfolio at time $t \in [0,T]$ is given by

$$V_t(\phi) = H_t^0 S_t^0 + H_t S_t. \tag{2.18}$$

The strategy ϕ will be called self-financing if

$$\mathrm{d}V_t(\phi) = H_t^0 \mathrm{d}S_t^0 + H_t \mathrm{d}S_t. \tag{2.19}$$

As both S and S^0 are Itô processes, we know that the above equation makes sense, provided

$$\int_0^T |H_t^0|\mathrm{d}t < \infty \text{ and } \int_0^T H_t^2\mathrm{d}t < \infty \;\; \mathbb{P}-\text{a.s.}$$

Remark 2.47. *Taking Definition 2.25 literally, we should require*

$$\begin{aligned}
&\int_0^T |rS_t^0 H_t^0|\mathrm{d}t = \int_0^T re^{rt}|H_t^0|\mathrm{d}t \leq re^{rT}\int_0^T |H_t^0|\mathrm{d}t < \infty,\\
&\int_0^T |\mu S_t H_t|\mathrm{d}t \leq |\mu| \left\|S\right\|_\infty \int_0^T |H_t|\mathrm{d}t < \infty, \text{ and}\\
&\int_0^T \sigma^2 S_t^2 H_t^2\mathrm{d}t \leq \sigma^2 \left\|S\right\|_\infty^2 \int_0^T H_t^2\mathrm{d}t < \infty,
\end{aligned}$$

where $\left\|S\right\|_\infty = \operatorname{esssup}\{|S_t| \mid t\in[0,T]\}$, which is finite because $t \mapsto S_t$ is continuous $\mathbb{P}$-a.s., which implies that S is bounded $\mathbb{P}$-a.s. on $[0,T]$. We see that these conditions are equivalent to the ones stated above. (Recall that $L^2 \subset L^1$ for finite measure spaces.)

Proposition 2.48. *Let S and S^0 be the price process described above and let $\widetilde{S}_t = e^{-rt}S_t$. Let $\phi = (H_t^0, H_t)_{t\in[0,T]}$ be an adapted process with values in $\mathbb{R}^2$ such that*

$$\int_0^T |H_t^0|\mathrm{d}t + \int_0^T H_t^2\mathrm{d}t < \infty \ \ \mathbb{P}-a.s.$$

Let V be defined as in (2.18) and let $\widetilde{V}_t(\phi) := e^{-rt}V_t(\phi)$. Then ϕ defines a self-financing strategy if and only if

$$\widetilde{V}_t(\phi) = V_0(\phi) + \int_0^t H_u\mathrm{d}\widetilde{S}_u, \quad \forall t \in [0,T]. \tag{2.20}$$

Proof. Let ϕ be a self-financing strategy. Then we have (from the product rule)

$$\mathrm{d}\widetilde{V}_t(\phi) = \mathrm{d}(e^{-rt}V_t(\phi)) = -re^{-rt}V_t(\phi)\mathrm{d}t + e^{-rt}\mathrm{d}V_t(\phi),$$

and similary

$$\mathrm{d}\widetilde{S}_t = \mathrm{d}(e^{-rt}S_t) = -r\widetilde{S}_t\mathrm{d}t + e^{-rt}\mathrm{d}S_t$$

hence from (2.19) and the definition of the value process V we infer that

$$\begin{aligned}
\mathrm{d}\widetilde{V}_t(\phi) &= -re^{-rt}(H_t^0S_t^0 + H_tS_t)\mathrm{d}t + e^{-rt}(H_t^0 \underbrace{\mathrm{d}S_t^0}_{=re^{rt}\mathrm{d}t} + H_t\mathrm{d}S_t) \\
&= -re^{-rt}(H_t^0e^{rt} + H_tS_t)\mathrm{d}t + rH_t^0\mathrm{d}t + H_t(\mathrm{d}\widetilde{S}_t + r\widetilde{S}_t\mathrm{d}t) \\
&= -rH_t\widetilde{S}_t\mathrm{d}t + H_t\mathrm{d}\widetilde{S}_t + rH_t\widetilde{S}_t\mathrm{d}t \\
&= H_t\mathrm{d}\widetilde{S}_t.
\end{aligned}$$

On the other hand, let (2.20) hold. Then, by the product rule,

$$H_t\mathrm{d}\widetilde{S}_t = \mathrm{d}\widetilde{V}_t(\phi) = \mathrm{d}(e^{-rt}V_t(\phi)) = -re^{-rt}V_t(\phi)\mathrm{d}t + e^{-rt}\mathrm{d}V_t(\phi),$$

hence by simple rearrangement we get

$$\begin{aligned}
\mathrm{d}V_t(\phi) &= e^{rt}H_t\mathrm{d}\widetilde{S}_t + rV_t(\phi)\mathrm{d}t \\
&= e^{rt}H_t(-r\widetilde{S}_t\mathrm{d}t + e^{-rt}\mathrm{d}S_t) + rV_t(\phi)\mathrm{d}t \\
&= H_t\mathrm{d}S_t + r(V_t(\phi) - H_tS_t)\mathrm{d}t \\
&= H_t\mathrm{d}S_t + rH_t^0S_t^0\mathrm{d}t \\
&= H_t\mathrm{d}S_t + H_t^0\mathrm{d}S_t^0,
\end{aligned}$$

which shows that ϕ is indeed self-financing. □

Once we have self-financing strategies, we would like to talk about equivalent martingale measures. To this end, we need to see what a measure change does to an Itô process (in its integral or differential form) and we want to gather conditions under which a process is a martingale under the new measure.
Before we can do this, however, we should introduce formally the notion of a stochastic differential equation (SDE) and what we consider as a solution of an SDE.

3 Stochastic Differential Equations and Diffusions

3.1 Introduction to SDEs

Let W be a standard BM on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $(\mathcal{F}_t)$ denote the filtration generated by W, augmented by the nullsets.
An equation of the form

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t \tag{3.1}$$

is called a *stochastic differential equation* (SDE) with (measurable) coefficients $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$.

Example 3.1. *For $\mu(t,x) = \mu \cdot x$ and $\sigma(t,x) = \sigma \cdot x$ we already know that the geometric BM,*

$$X_t = X_0 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right),$$

satisfies the corresponding SDE

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t. \tag{3.2}$$

Example 3.2 (Ornstein-Uhlenbeck process / Langevin equation)**.** *For constants $\alpha, \sigma > 0$ and $x_0 \in \mathbb{R}$ consider the SDE*

$$\mathrm{d}X_t = -\alpha X_t \mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0 = x_0. \tag{3.3}$$

First observe that the drift term $-\alpha X_t$ is positive whenever X_t is negative and it is negative whenever X_t is positive. Therefore, the process can be expected to fluctuate around zero.
In order to get a first intuition, let $\sigma = 0$. Then the resulting ODE has the solution $X_t^{\sigma=0} = x_0 e^{-\alpha t}$, hence $X_t^{\sigma=0} \cdot e^{\alpha t}$ is constant.
Now consider $Y_t = X_t \cdot e^{\alpha t}$ for a solution X of (3.3)*. In particular, $Y_0 = X_0 \cdot 1 = x_0$. By applying the product rule and plugging in* (3.3)*, we get*

$$\begin{aligned}\mathrm{d}Y_t &= e^{\alpha t}\mathrm{d}X_t + \alpha e^{\alpha t} X_t \mathrm{d}t \\ &= e^{\alpha t}(-\alpha X_t \mathrm{d}t + \sigma \mathrm{d}W_t) + \alpha e^{\alpha t} X_t \mathrm{d}t \\ &= \sigma e^{\alpha t}\mathrm{d}W_t,\end{aligned}$$

hence $Y_t = x_0 + \int_0^t \sigma e^{\alpha s}\mathrm{d}W_s$ for any $t \in [0,T]$. Consequently,

$$X_t = e^{-\alpha t} Y_t = e^{-\alpha t}\left(x_0 + \int_0^t \sigma e^{\alpha s}\mathrm{d}W_s\right). \tag{3.4}$$

The process X in (3.4) *is called* Ornstein-Uhlenbeck process.

In both examples we have found process that satisfy the given SDE. However, we do not know yet whether the solution we found is unique.
The following result (in the spirit of the Picard-Lindelöf theorem) gives sufficient conditions such that an SDE with initial value has a solution and that said solution is unique in a certain sense.

Theorem 3.3. *For $T > 0$ and measurable functions $\mu, \sigma \colon [0,T] \times \mathbb{R} \to \mathbb{R}$, consider the SDE with initial value*

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \quad t \in [0,T], \quad X_0 = Z. \tag{3.5}$$

Make the following assumptions:

(i) Z is independent of $\sigma(W_t, t \geq 0)$ and Z is square integrable, i.e., that $\mathbb{E}[|Z|^2] < \infty$.

(ii) There exists a constant $C \geq 0$ sucht that for any $x \in \mathbb{R}$ and $t \in [0,T]$,

$$|\mu(t,x)| + |\sigma(t,x)| \leq C(1 + |x|).$$

(iii) There exists a constant $D \geq 0$ such that for any $x, y \in \mathbb{R}$ and $t \in [0,T]$, the Lipschitz condition

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x - y|$$

is satisfied.

Then there exists a unique continuous process X such that

(i) X is adapted to the filtration $(\mathcal{F}_t^Z)$ generated by Z and $(W_s)_{s \leq t}$,

(ii) X satisfies

$$X_t = Z + \int_0^t \mu(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s,$$

(iii) $\mathbb{E}\left[\int_0^T |X_t|^2 \mathrm{d}t\right] < \infty$.

Let us make a few remarks:

Remark 3.4.

1. *A solution with the above properties is called a* strong *solution to the SDE. For a given BM W, it satisfies the SDE. In contrast, a weak solution is one where there exist a filtration, a BM B (under the filtration) and a process X such that X is adapted to the filtration an satisfies the SDE with B.*

2. *In [SKY01], quadratic conditions replace (ii) and (iii) on μ and σ. Consequently, they obtain a solution X that is uniformly bounded in $L^2(\mathbb{P})$, i.e.,*

$$\sup\left\{ \mathbb{E}[X_t^2] \,\middle|\, t \in [0,T] \right\} < \infty.$$

3. *The SDE for the geometric BM,* (3.2)*, with $\mu(t,x) = \mu \cdot x$ and $\sigma(t,x) = \sigma \cdot x$ satisfies the conditions with $C = D = |\mu| + |\sigma|$, hence our solution is indeed unique.*

4. *The Langevin equation (for the Ornstein-Uhlenbeck process)* (3.3) *with $\mu(t,x) = -\alpha \cdot x$ and $\sigma(t,x) = \sigma$ satisfies the conditions of the theorem as well. (Check this as a quick exercise!) Therefore, the Ornstein-Uhlenbeck process is indeed the only process satisfying* (3.3).

Example 3.5 (A counterexample). *Let us have a look at the* Tanaka equation

$$\mathrm{d}X_t = \operatorname{sgn}(X_t)\mathrm{d}W_t, \quad \textit{where } \operatorname{sgn}(x) = \begin{cases} +1, & \textit{if } x \geq 0, \\ -1, & \textit{if } x < 0. \end{cases} \tag{3.6}$$

The Lipschitz condition is not satisfied by $\mu(t,x) = 0$ and $\sigma(t,x) = \operatorname{sgn}(x)$. Let B be a standard BM and consider the process

$$Y_t = \int_0^t \frac{1}{\operatorname{sgn}(B_s)} \mathrm{d}B_s = \int_0^t \operatorname{sgn}(B_s)\mathrm{d}B_s.$$

The process $\operatorname{sgn}(B_s)$ is adapted to the filtration generated by B_s and $\int_0^T (\operatorname{sgn} B_s)^2 \mathrm{d}t = T < \infty$, hence Y is well defined. Its quadratic variation is

$$\langle Y, Y \rangle_t = \int_0^t (\operatorname{sgn} B_s)^2 \mathrm{d}\langle B \rangle_s = \int_0^t 1 \mathrm{d}s = t.$$

By Lévy's characterization of BM[8]*, Y is a BM satisfying (from its first definition)*

$$\mathrm{d}Y_t = \frac{\mathrm{d}B_t}{\operatorname{sgn}(B_t)} \qquad \Longleftrightarrow \qquad \mathrm{d}B_t = \operatorname{sgn}(B_t)\mathrm{d}Y_t,$$

hence B solves (3.6) *with BM Y. It is, however, only a weak (and not a strong) solution of* (3.6)*. To see this, let $(\mathcal{F}_t^B)$ be the filtration generated by B and let $(\mathcal{F}_t^Y)$ be the filtration generated by Y. Then $\mathcal{F}^Y \subsetneq \mathcal{F}^B$; in particular, $\operatorname{sgn}(B_t)$ is not $(\mathcal{F}_t^Y)$-adapted. One way to prove is is to apply Tanaka's formula, which we will not discuss in this lecture. Alternatively, consider a sequence $(f_n)_{n \in \mathbb{N}} \subset C^1(\mathbb{R})$ such that*

- *$f_n(x) = \operatorname{sgn}(x)$ for $|x| \geq \frac{1}{n}$,*
- *$|f_n(x)| \leq 1$ and $f_n(-x) = -f_n(x)$ for all $x \in \mathbb{R}$.*

Then $F_n(x) := \int_0^x f_n(y)\mathrm{d}y \in C^2(\mathbb{R})$ and $\lim_{n \to \infty} F_n(x) = |x|$ uniformly on compact intervals. By Itô's formula,

$$F_n(B_t) - \int_0^t f_n(B_s)\mathrm{d}B_s = \frac{1}{2}\int_0^t f_n'(B_s)\mathrm{d}s = \frac{1}{2}\int_0^t f_n'(|B_s|)\mathrm{d}s$$

For $n \to \infty$ we therefore get

$$|B_t| - \int_0^t \operatorname{sgn}(B_s)\mathrm{d}B_s = |B_t| - Y_t = \textit{something } \mathcal{F}_t^{|B|}\textit{-measurable},$$

hence Y_t is $\mathcal{F}_t^{|B|}$-measurable. In other words, $\mathcal{F}_t^Y \subset \mathcal{F}_t^{|B|}$ (for all $t \geq 0$), and the latter is a strict subset of $\mathcal{F}_t^B$.

[8] Lévy's characterization of the one-dimensional BM says that any continuous local martingale M with $M_0 = 0$ is a BM iff $\langle M \rangle_t = t$.

3.2 Solutions to linear SDEs

So far all SDEs had a closed form solution. In this section we look at a class of SDEs for which this is also true: linear SDEs, i.e., SDEs of the form

$$\mathrm{d}X_t = (\alpha_t + \beta_t X_t)\mathrm{d}t + (\gamma_t + \delta_t X_t)\mathrm{d}W_t. \tag{3.7}$$

The special case $\alpha \equiv \delta \equiv 0$ leads to the Ornstein-Uhlenbeck process that we have just seen in the last section; the case $\alpha \equiv \gamma \equiv 0$ leads to a stochastic exponential with the special case of a geometric BM for $\beta_t \equiv \mu$ and $\delta_t \equiv \sigma$.
Let us first introduce the stochastic exponential in a more general form than (2.13).

3.2.1 Stochastic Exponential

Consider the linear SDE

$$\mathrm{d}U_t = \beta_t U_t \mathrm{d}t + \delta_t U_t \mathrm{d}W_t. \tag{3.8}$$

If Y is an Itô process with

$$\mathrm{d}Y_t = \beta_t \mathrm{d}t + \delta_t \mathrm{d}W_t,$$

then (3.8) can be rewritten as

$$\mathrm{d}U_t = U_t \mathrm{d}Y_t.$$

Definition 3.6. *A process U satisfying* (3.8) *is called the* stochastic exponential *of Y. It is denoted by*

$$U_t = U_0 \mathcal{E}(Y)_t.$$

Lemma 3.7. *The stochastic exponential of an Itô process Y with $Y_0 = 0$ is given by*

$$\mathcal{E}(Y)_t = \exp\left(Y_t - \frac{1}{2}\langle Y\rangle_t\right).$$

Proof. Let $U_t = \mathcal{E}(Y)_t$. Without loss of generality let $U_0 = 1$. On $\{U_t > 0\}$ (which is always the case for $U_0 > 0$) we have by Itô's formula

$$\mathrm{d}(\ln U_t) = \frac{1}{U_t}\mathrm{d}U_t - \frac{1}{2}\frac{1}{U_t^2}\mathrm{d}\langle U\rangle_t = \mathrm{d}Y_t - \frac{1}{2}\mathrm{d}\langle Y\rangle_t,$$

hence

$$\ln \mathcal{E}(Y)_t = Y_t - \frac{1}{2}\langle Y\rangle_t,$$

which proves the statement. □

If we replace Y by its representation as an Itô process and recall that $\langle Y\rangle_t = \int_0^t \delta_s^2 \mathrm{d}s$, we get the representation for U:

$$U_t = U_0 \exp\left(\int_0^t \left(\beta_s - \frac{1}{2}\delta_s^2\right)\mathrm{d}s + \int_0^t \delta_s \mathrm{d}W_s\right) \tag{3.9}$$

Application in Finance

If the price of a given stock is given by an Itô process S and the return is denoted by R, then

$$\mathrm{d}R_t = \frac{\mathrm{d}S_t}{S_t} \quad \text{or equivalently} \quad \mathrm{d}S_t = S_t \mathrm{d}R_t.$$

Hence, the stock price is the stochastic exponential of the return, i.e., $S_t = S_0 \mathcal{E}(R)_t$.

Stochastic Logarithm

If $U = \mathcal{E}(Y)$, then Y is called the *stochastic logarithm* of U and is denoted by $Y = \mathcal{L}(U)$. From our previous calculations we can infer the following properties of the stochastic logarithm:

Corollary 3.8. *Let $U > 0$. Then the stochastic logarithm, if it exists, satisfies the SDE*

$$\mathrm{d}Y_t = \frac{\mathrm{d}U_t}{U_t}, \quad Y_0 = 0.$$

It has the representation

$$Y_t = \mathcal{L}(U)_t = \ln\left(\frac{U_t}{U_0}\right) + \int_0^t \frac{1}{2U_s^2} \mathrm{d}\langle U \rangle_s.$$

Example 3.9. *Let $U_t = e^{W_t}$ for a standard BM W. It satisfies*

$$\mathrm{d}U_t = e^{W_t} \mathrm{d}W_t + \frac{1}{2} e^{W_t} \mathrm{d}t.$$

Its stochastic logarithm therefore satisfies

$$\mathrm{d}Y_t = \mathrm{d}\mathcal{L}(U)_t = \frac{\mathrm{d}U_t}{U_t} = \mathrm{d}W_t + \frac{1}{2}\mathrm{d}t,$$

hence

$$\mathcal{L}(U)_t = W_t + \frac{1}{2}t.$$

Let us verify the representation asserted in the lemma:

$$\ln\left(\frac{U_t}{U_0}\right) + \int_0^t \frac{1}{2U_s^2} \mathrm{d}\langle U \rangle_s = \ln(e^{W_t}) + \int_0^t \frac{1}{2e^{2W_s}} \cdot e^{2W_s} \mathrm{d}s = W_t + \frac{1}{2}t,$$

where we used that

$$\mathrm{d}\langle U \rangle_t = \mathrm{d}U_t \mathrm{d}U_t = e^{2W_t} \mathrm{d}t.$$

3.2.2 General linear SDEs

We will use a trick known from ODEs, namely variation of constants. As before, let U satisfy (3.8), i.e.,

$$\mathrm{d}U_t = \beta_t U_t \mathrm{d}t + \delta_t U_t \mathrm{d}W_t.$$

Furthermore, let V satisfy

$$\mathrm{d}V_t = a_t \mathrm{d}t + b_t \mathrm{d}W_t$$

and impose the initial condition $U_0 = 1$ and $V_0 = Z$. We want to find a solution to the general linear SDE (3.7) with initial condition $X_0 = Z$. If we consider the differential of the product, $\mathrm{d}X_t = \mathrm{d}(UV)_t$, we get

$$\begin{aligned}
\mathrm{d}X_t &= V_t \mathrm{d}U_t + U_t \mathrm{d}V_t + \mathrm{d}\langle U, V\rangle_t \\
&= V_t(\beta_t U_t \mathrm{d}t + \delta_t U_t \mathrm{d}W_t) + U_t(a_t \mathrm{d}t + b_t \mathrm{d}W_t) + (\beta_t U_t \mathrm{d}t + \delta_t U_t \mathrm{d}W_t) \cdot (a_t \mathrm{d}t + b_t \mathrm{d}W_t) \\
&= \beta_t X_t \mathrm{d}t + \delta_t X_t \mathrm{d}W_t + a_t U_t \mathrm{d}t + b_t U_t \mathrm{d}W_t + b_t \delta_t U_t \mathrm{d}t \\
&\overset{!}{=} (\alpha_t + \beta_t X_t)\mathrm{d}t + (\gamma_t + \delta_t X_t)\mathrm{d}W_t.
\end{aligned}$$

A comparison of coefficients tells us that

$$a_t U_t + b_t \delta_t U_t \overset{!}{=} \alpha_t \quad \text{und} \quad b_t U_t \overset{!}{=} \gamma_t.$$

Replacing $b_t U_t$ by γ_t in the first condition and rearrangement gives the equivalent set of conditions

$$a_t U_t \overset{!}{=} \alpha_t - \gamma_t \delta_t \quad \text{und} \quad b_t U_t \overset{!}{=} \gamma_t.$$

Recall that in Equation (3.9) we had the explicit expression

$$U_t = U_0 \exp\left(\int_0^t \left(\beta_s - \frac{1}{2}\delta_s^2\right) \mathrm{d}s + \int_0^t \delta_s \mathrm{d}W_s\right)$$

for U_t, which gives the following representation of X:

$$\begin{aligned}
X_t = U_t V_t &= U_t \left(Z + \int_0^t a_s \mathrm{d}s + \int_0^t b_s \mathrm{d}W_s\right) \\
&= U_t \left(Z + \int_0^t \frac{\alpha_s - \gamma_s \delta_s}{U_s} \mathrm{d}s + \int_0^t \frac{\gamma_s}{U_s} \mathrm{d}W_s\right).
\end{aligned}$$

Example 3.10 (Langevin type SDE)**.** *Consider the linear SDE*

$$\mathrm{d}X_t = \beta_t X_t \mathrm{d}t + \mathrm{d}W_t, \tag{3.10}$$

i.e., $\alpha \equiv \delta \equiv 0$ *and* $\gamma \equiv 1$*. Let us first solve for the corresponding process U, which satisfies*

$$\mathrm{d}U_t = \beta_t U_t \mathrm{d}t, \quad U_0 = 1.$$

The unique solution to this equation is given by $U_t = \exp\left(\int_0^t \beta_s \mathrm{d}s\right)$*. Therefore, the solution* X *to* (3.10) *is*

$$X_t = U_t\left(X_0 + \int_0^t \frac{1}{U_s}\mathrm{d}W_s\right) = \exp\left(\int_0^t \beta_s \mathrm{d}s\right)\left(X_0 + \int_0^t \exp\left(-\int_0^s \beta_r \mathrm{d}r\right) \mathrm{d}W_s\right).$$

Example 3.11 (Brownian bridge). *Consider the linear SDE*

$$\mathrm{d}X_t = \frac{b - X_t}{T-t}\mathrm{d}t + \mathrm{d}W_t,\ 0 \le t < T, \quad X_0 = a. \tag{3.11}$$

The coefficients in the context of the general linear SDE are

$$\alpha_t = \frac{b}{T-t}, \quad \beta_t = \frac{-1}{T-t}, \quad \gamma_t \equiv 1, \quad \delta_t \equiv 0.$$

As before, we first identify U as the solution of the SDE

$$\mathrm{d}U_t = \beta_t U_t \mathrm{d}t + \delta_t U_t \mathrm{d}W_t = \frac{-U_t}{T-t}\mathrm{d}t,$$

hence

$$U_t = \exp\left(\int_0^t \frac{-1}{T-s}\mathrm{d}s\right) = \exp\left(\ln(T-t) - \ln(T-0)\right) = \frac{T-t}{T}.$$

Consequently, X is given by

$$\begin{aligned}
X_t &= U_t\left(a + \int_0^t \frac{\alpha_s}{U_s}\mathrm{d}s + \int_0^t \frac{1}{U_s}\mathrm{d}W_s\right)\\
&= \left(\frac{T-t}{T}\right)\cdot\left(a + \int_0^t \frac{bT}{(T-s)^2}\mathrm{d}s + \int_0^t \frac{T}{T-s}\mathrm{d}W_s\right)\\
&= \left(\frac{T-t}{T}\right)\cdot\left(a + \frac{bT}{T-t} - \frac{bT}{T-0} + \int_0^t \frac{T}{T-s}\mathrm{d}W_s\right)\\
&= a\left(1 - \frac{t}{T}\right) + b\frac{t}{T} + (T-t)\int_0^t \frac{1}{T-s}\mathrm{d}W_s.
\end{aligned}$$

Observe that for any $t < T$, $\int_0^t \left(\frac{1}{T-s}\right)^2 \mathrm{d}s = \frac{1}{T-t} - \frac{1}{T} < \infty$. The integrand of the stochastic integral is therefore measurable (because deterministic) and square integrable, hence it belongs to $\mathcal{V}$. The process $\int_0^t \frac{1}{T-s}\mathrm{d}W_s$ is therefore a (continuous) martingale and a Gaussian process. If one verifies that $X_T = b$, then from the representation formula for X we can infer that $\mathbb{E}[X_t] = a\left(1 - \frac{t}{T}\right) + b\frac{t}{T}$ and $\mathrm{Cov}(X_t, X_s) = \min(s,t) - \frac{st}{T}$.

Exercise 10. *Do the following to complete the above example:*

a) *Show that $\lim_{t\uparrow T}(T-t)\int_0^t \frac{1}{T-s}\mathrm{d}W_s = 0$ almost surely. This implies that the Brownian Bridge X has fixed initial and terminal values with $X_0 = a$ and $X_T = b$ almost surely.*

b) *Verify the expected value and covariance of X.*

Before we continue with the next topic, let us remark why we added the term *diffusions* in this chapter's title. The reason is that equations of the form

$$X_t = X_0 + \int_0^t \mu(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s$$

are called *diffusion-type SDEs* and strong solutions to the corresponding SDEs are called *Itô diffusions*. The name originates from the corresponding physical phenomena which can be described in this manner, e.g. the famous motion of a particle in a fluid.

3.3 Change of measures

In FiMa I we have seen that martingale measures, i.e. probability measures under which a share's price equals its discounted expected value, play an important role in asset pricing. In particular, in a discrete market the fundamental theorem of asset pricing (FTAP) tells us that the existence of an equivalent martingale measure (EMM) is equivalent to the non-existence of arbitrage.
Before we start with risk neutral pricing, let us recall what we already know about change of measures and extend that knowledge for models in continuous time. We will find the stochastic exponential quite useful in this section.

Setting

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space; the filtration $(\mathcal{F}_t)_{t\in[0,T]}$ is assumed to satisfy the usual conditions; w.l.o.g. $\mathcal{F} = \mathcal{F}_T$.
Recall that if a probability measure $\mathbb{Q}$ is absolutely continuous w.r.t. $\mathbb{P}$ on $\mathcal{F}$ (i.e. $\mathbb{Q} \ll \mathbb{P}$), then $\mathbb{Q}|_{\mathcal{F}_t} \ll \mathbb{P}|_{\mathcal{F}_t}$ and there exists the Radon-Nikodym derivative $Z := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ such that $Z_t := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t}$ satisfies

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[Z_T \mathbb{1}_A], \quad \forall A \in \mathcal{F}_T.$$

Z is a uniformly integrable martingale with $\mathbb{E}^{\mathbb{P}}[Z_T] = 1$. It is chosen to be càdlàg.
Recall Bayes' formula:

Lemma 3.12 (Bayes). *Let $0 \le s \le t \le T$. Assume that $Z_t > 0$ for all t (and in particular $Z > 0$). Then $\mathbb{P} \sim \mathbb{Q}$ and we have*

$$\mathbb{E}^{\mathbb{Q}}[Y|\mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}[YZ|\mathcal{F}_s]}{\mathbb{E}^{\mathbb{P}}[Z|\mathcal{F}_s]}, \quad \text{and} \quad \mathbb{E}^{\mathbb{Q}}[Y|\mathcal{F}_s] = \frac{\mathbb{E}^{\mathbb{P}}[YZ_t|\mathcal{F}_s]}{Z_s} \quad \text{if } Y \text{ is } \mathcal{F}_t\text{-mble,}$$

provided $Y \ge 0$ or $YZ \in L^1(\mathbb{P})$.

The following result links local martingales under $\mathbb{Q}$ and $\mathbb{P}$.

Lemma 3.13. *Let $R := \inf\{t \ge 0 \mid Z_t = 0\}$. Then*

(i) $R = \infty$ $\mathbb{Q}$-a.s.;

(ii) for a non-negative adapted process U or $UZ \in L^1(\mathbb{P})$, and $0 \le s < t$,

$$\mathbb{E}^{\mathbb{Q}}[U_t|\mathcal{F}_s] = \mathbb{1}_{\{Z_s \ne 0\}} \frac{1}{Z_s} \mathbb{E}^{\mathbb{P}}[U_t Z_t|\mathcal{F}_s] \quad \mathbb{Q}-a.s.;$$

(iii) if Y is an adapted process s.t. YZ is a local martingale under $\mathbb{P}$ with $Y_0 Z_0 \in L^1(\mathbb{P})$, then Y is a local martingale under $\mathbb{Q}$.

Proof. (i) The Radon-Nikodym derivative $Z = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ was chosen to be right-continuous, hence $Z_R \equiv 0$ on $\{R < \infty\}$. Therefore, for $t \ge 0$,

$$\mathbb{Q}(R \le t) = \mathbb{E}^{\mathbb{P}}[Z_t \mathbb{1}_{\{R\le t\}}] \overset{(*)}{=} \mathbb{E}^{\mathbb{P}}[Z_{t\wedge R} \mathbb{1}_{\{R\le t\}}] = \mathbb{E}^{\mathbb{P}}[Z_R \mathbb{1}_{\{R\le t\}}] = 0,$$

where $(*)$ holds by a stopping theorem because Z is uniformly integrable.

(ii) For $s < t$ let $A \in \mathcal{F}_s$. Then, as $Z_t > 0$ $\mathbb{Q}$-a.s. by (i),

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[U_t \mathbb{1}_A\right] &\overset{(i)}{=} \mathbb{E}^{\mathbb{Q}}\left[U_t \mathbb{1}_{A\cap\{Z_s\neq 0\}}\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[Z_t U_t \mathbb{1}_{A\cap\{Z_s\neq 0\}}\right] \\
&= \mathbb{E}^{\mathbb{P}}\left[\frac{\mathbb{E}^{\mathbb{P}}\left[U_t Z_t|\mathcal{F}_s\right]}{Z_s} Z_s \mathbb{1}_{A\cap\{Z_s\neq 0\}}\right] \\
&\overset{\text{Bayes}}{=} \mathbb{E}^{\mathbb{Q}}\left[\frac{\mathbb{E}^{\mathbb{P}}\left[U_t Z_t|\mathcal{F}_s\right]}{Z_s} \mathbb{1}_A \mathbb{1}_{\{Z_s\neq 0\}}\right],
\end{aligned}$$

where we used $\mathbb{E}^{\mathbb{P}}[Z_s] = 1$ in the application of Bayes' formula.[9]

(iii) Let (τ_n) be a localizing sequence for (YZ) such that the stopped process $(YZ - Y_0 Z_0)^{\tau_n}$ is a $\mathbb{P}$-martingale.[10]

First, let us check the integrability of Y^{τ_n} for $n \in \mathbb{N}$.

- $\mathbb{E}^{\mathbb{Q}}\left[Y^+_{t\wedge\tau_n}\right] = \mathbb{E}^{\mathbb{P}}\left[Z_{t\wedge\tau_n} Y^+_{t\wedge\tau_n}\right]$
- $\mathbb{E}^{\mathbb{Q}}\left[Y^-_{t\wedge\tau_n}\right] = \mathbb{E}^{\mathbb{P}}\left[Z_{t\wedge\tau_n} Y^-_{t\wedge\tau_n}\right]$
- $\mathbb{E}^{\mathbb{P}}\left[Z_{t\wedge\tau_n} Y_{t\wedge\tau_n}\right] = \mathbb{E}^{\mathbb{P}}\left[Z_0 Y_0\right]$ because $(YZ)^{\tau_n}$ is a martingale

From the above points we can infer that $Y_{t\wedge\tau_n} \in L^1(\mathbb{Q})$ for all $t \geq 0$ and all $n \in \mathbb{N}$.

Now let us check the martingale property of $Y_{t\wedge\tau_n}$ w.r.t. $\mathbb{Q}$. Let $0 \leq s < t$. Then

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[Y_{t\wedge\tau_n}|\mathcal{F}_s\right] &= \mathbb{1}_{\{\tau_n\leq s\}} Y_{\tau_n} + \mathbb{1}_{\{\tau_n > s\}} \mathbb{E}^{\mathbb{Q}}\left[Y_{t\wedge\tau_n}|\mathcal{F}_s\right] \\
&\overset{(ii)}{=} \mathbb{1}_{\{\tau_n\leq s\}} Y_{\tau_n} + \mathbb{1}_{\{\tau_n > s\}} \mathbb{1}_{\{Z_s\neq 0\}} \frac{\mathbb{E}^{\mathbb{P}}\left[Y_{t\wedge\tau_n} Z_{t\wedge\tau_n}|\mathcal{F}_s\right]}{Z_s} \\
&\overset{\text{mart.}}{=} \mathbb{1}_{\{\tau_n\leq s\}} Y_{\tau_n} + \mathbb{1}_{\{\tau_n > s\}} \mathbb{1}_{\{Z_s\neq 0\}} \frac{Y_s Z_s}{Z_s} \\
&\overset{(i)}{=} Y_{s\wedge\tau_n} \quad \mathbb{Q}-\text{a.s.}
\end{aligned}$$

The martingale property of $(Y_{t\wedge\tau_n} - Y_0)$ follows directly, hence Y is a local martingale.

□

Let us recall in a slightly modified version Lévy's characterization of a standard BM, which we will need for the proof of Girsanov's theorem.

Theorem 3.14. *For a continuous real-valued process X with $X_0 = 0$ the following properties are equivalent:*

(i) X is a BM;

(ii) X is a local martingale and $\langle X\rangle_t = t$ a.s. for all t;

(iii) X is a local martingale and $X_t^2 - t$ is one as well.

[9] One can save the second and third line by replacing U_t by $\mathbb{E}^{\mathbb{Q}}\left[U_t|\mathcal{F}_s\right]$ in the first line (by the tower property) and applying Bayes directly to $\mathbb{E}^{\mathbb{Q}}\left[U_t|\mathcal{F}_s\right]$ with the $\mathcal{F}_t$-measurable random variable U_t.

[10] We write $X^\tau = (X_{t\wedge\tau})$ for the stopped process.

Proof. **(i) ⇒ (iii)** If X is a BM, then $\langle X\rangle_t = t$ for all t and by Itô's formula, $W_t^2 - t = 2\int_0^t W_s \mathrm{d}W_s$ (holds for X in place of W) and any stochastic integral (for an integrand in $\mathcal{W}$) is a local martingale.

(iii) ⇒ (ii) We have

$$X_t^2 - t = 2\int_0^t X_s \mathrm{d}X_s + \langle X\rangle_t - t.$$

The LHS is a local martingale, the stochastic integral is one. Consequently, also $Y_t := \langle X\rangle_t - t$ is a local martingale. (The sum of local martingales is also a local martingale.)

Recall that the quadratic variation of a continuous local martingale is continuous and nondecreasing, hence it is of bounded variation. The quadratic variation of the (continuous) quadratic variation therefore vanishes. In our setting, this implies $\langle Y\rangle_t = 0$, hence

$$Y_t^2 - \langle Y\rangle_t = Y_t^2 = 2\int_0^t Y_s \mathrm{d}Y_s,$$

is a local martingale. (Verify!) Thus, there exists a localizing sequence (τ_n) such that

$$\mathbb{E}\left[Y_{t\wedge\tau_n}^2\right] = 0, \quad \forall t \in [0,T], \forall n \in \mathbb{N}.$$

Hence, $Y_{t\wedge\tau_n} = 0$ almost surely; by passing to the limit ($n \to \infty$) we get that $Y_t = 0$ almost surely, which proves the claim.

(ii) ⇒ (i) Let $u \in \mathbb{R}$ and $0 \leq s \leq t \leq T$. Then

$$\begin{aligned}
\mathbb{E}\left[e^{iu(X_t - X_s)} \,\middle|\, \mathcal{F}_s\right] &= \mathbb{E}\left[\frac{\mathcal{E}(iuX)_t}{\mathcal{E}(iuX)_s} e^{-\frac{1}{2}u^2(\langle X\rangle_t - \langle X\rangle_s)} \,\middle|\, \mathcal{F}_s\right] \\
&= e^{-\frac{1}{2}u^2(t-s)}\mathbb{E}\left[\frac{\mathcal{E}(iuX)_t}{\mathcal{E}(iuX)_s} \,\middle|\, \mathcal{F}_s\right] \\
&= e^{-\frac{1}{2}u^2(t-s)},
\end{aligned}$$

where we used that the stochastic exponential of a local martingale is itself a local martingale and a bounded local martingale (such as $\mathcal{E}(iuX)$) is even a martingale.

Knowing the characteristic function of $X_t - X_s$, we can infer that $X_t - X_s \sim \mathcal{N}(0, t-s)$. In particular, $X_t - X_s$ is independent of $\mathcal{F}_s$ with that distribution. Hence, X is a BM.

□

Theorem 3.15 (Girsanov Theorem I)**.** *Let $\alpha = (\alpha_t) \in \mathcal{W}$ and let Y be the solution to the SDE*

$$\mathrm{d}Y_t = \alpha_t \mathrm{d}t + \mathrm{d}W_t, \quad t \in [0,T], \quad Y_0 = 0.$$

Let $M = \mathcal{E}(\int -\alpha \mathrm{d}W)$, i.e.,

$$M_t = \exp\left(-\int_0^t \alpha_s \mathrm{d}W_s - \frac{1}{2}\int_0^t \alpha_s^2 \mathrm{d}s\right), \quad t \in [0,T].$$

If M is a martingale w.r.t. $(\mathcal{F}_t)$ and $\mathbb{P}$, then Y is a standard BM w.r.t. the equivalent measure $\mathbb{Q} \sim \mathbb{P}$ defined by

$$\mathrm{d}\mathbb{Q} = M_T \mathrm{d}\mathbb{P}.$$

Before we prove this theorem, let us remark that we will indeed have a second theorem of Girsanov. The difference is that in the above theorem, the coefficient of the stochastic integral is fixed to be 1, whereas we will later on relax this requirement.

Proof. First, let us observe that

$$\mathbb{Q}(\Omega) = \mathbb{E}^{\mathbb{P}}[M_T] \overset{(*)}{=} \mathbb{E}^{\mathbb{P}}[M_0] = 1,$$

where $(*)$ holds because M is a $\mathbb{P}$-martingale and the last equality holds because $M_0 = 1$. This shows that $\mathbb{Q}$ is indeed a probability measure.
Furthermore, $\mathbb{Q} \sim \mathbb{P}$ (i.e. the measures are equivalent), because $M > 0$ $\mathbb{P}$-almost surely. Consider the process $K_t := M_t Y_t$ ($t \in [0, T]$). We know that M satisfies the SDE

$$\mathrm{d}M_t = -\alpha_t M_t \mathrm{d}W_t,$$

hence with the product rule we find that

$$\begin{aligned}
\mathrm{d}K_t &= M_t \mathrm{d}Y_t + Y_t \mathrm{d}M_t + \mathrm{d}\langle M, Y\rangle_t \\
&= M_t(\alpha_t \mathrm{d}t + \mathrm{d}W_t) - Y_t \alpha_t M_t \mathrm{d}W_t - \alpha_t M_t \mathrm{d}t \\
&= M_t(1 - \alpha_t Y_t)\mathrm{d}W_t.
\end{aligned}$$

Taking into consideration that $\mathbb{E}^{\mathbb{P}}[K_0] = \mathbb{E}^{\mathbb{P}}[M_0 Y_0] = 0$ (because $Y_0 = 0$ $\mathbb{P}$-a.s.), we can infer that K is a local martingale under $\mathbb{P}$. Consequently, by Lemma 3.13, Y is a local martingale under $\mathbb{Q}$. Similarly, one can show that $(M_t(Y_t^2 - t))_t$ is a local martingale w.r.t. $\mathbb{P}$, which implies that $Y_t^2 - t$ is $\mathbb{Q}$-local martingale. Hence, by Lévy's characterization (Theorem 3.14), Y is a BM.[11] □

Caution! Theorem 3.15 requires that the stochastic exponential M is a martingale — not only a local martingale. We will present some useful sufficient conditions that guarantee that the stochastic exponential is indeed a martingale.

Sufficient conditions

Recall the notation

$$\mathcal{E}(L)_t = \exp\left(L_t - \frac{1}{2}\langle L\rangle_t\right).$$

Theorem 3.16 (Kazamaki's condition)**.** *Let L be a local martingale with $L_0 = 0$. Let $\mathcal{T}_T := \{\tau \text{ stopping time} \mid \tau \le T\}$. If*

$$\sup_{\tau \in \mathcal{T}_T} \mathbb{E}\left[e^{\frac{1}{2}L_\tau}\right] < \infty,$$

then $M = \mathcal{E}(L)$ is a martingale on $[0, T]$.

Before we can prove this result, we need an auxiliary one:

Lemma 3.17. *Let $1 < p < \infty <$ and $\frac{1}{p} + \frac{1}{q} = 1$. Assume that L is a local martingale on $[0, T]$ with $L_0 = 0$. Assume furthermore that*

$$\sup_{\tau \in \mathcal{T}_T} \mathbb{E}\left[\exp\left(\frac{\sqrt{p}}{2(\sqrt{p} - 1)} L_\tau\right)\right] < \infty.$$

Then $M := \mathcal{E}(L)$ is an (L^q-bounded) martingale.

[11] Alternatively one could have calculated $\langle Y\rangle_t = \langle W\rangle_t = t$ to get the other characterization of a BM stated in Theorem 3.14.

Proof of Lemma 3.17.
Let $\tau \in \mathcal{T}_T$. Set $u := \frac{\sqrt{p}+1}{\sqrt{p}-1}$ and $v := \frac{\sqrt{p}+1}{2}$. Then $u > 1$, $\frac{1}{u} + \frac{1}{v} = 1$ and with $q = \frac{p}{p-1}$ we have

$$\begin{aligned}\left(q - \sqrt{\frac{q}{u}}\right) v &= \frac{1}{2}(\sqrt{p}+1)\left(\frac{p}{p-1} - \sqrt{\frac{p(\sqrt{p}-1)}{(p-1)(\sqrt{p}+1)}}\right) \\ &= \frac{\sqrt{p}}{2}(\sqrt{p}+1)\left(\frac{\sqrt{p}}{p-1} - \frac{1}{\sqrt{p}+1}\right) \\ &= \frac{\sqrt{p}}{2(\sqrt{p}-1)}(p-1) \cdot \frac{\sqrt{p} - ((\sqrt{p}-1))}{p-1} \\ &= \frac{\sqrt{p}}{2(\sqrt{p}-1)}.\end{aligned} \tag{3.12}$$

Recall Hölder's inequality for random variables X and Y:

$$\mathbb{E}\left[|XY|\right] \leq \left(\mathbb{E}\left[|X|^u\right]\right)^{\frac{1}{u}} \cdot \left(\mathbb{E}\left[|Y|^v\right]\right)^{\frac{1}{v}} \quad \text{if } \frac{1}{u} + \frac{1}{v} = 1.$$

With the multiplicative representation

$$(\mathcal{E}(L)_\tau)^q = e^{qL_\tau - \frac{q}{2}\langle L\rangle_\tau} = e^{\sqrt{\frac{q}{u}}L_\tau - \frac{q}{2}L_\tau} \cdot e^{\left(q - \sqrt{\frac{q}{u}}\right)L_\tau}$$

we can apply Hölder's inequality to get

$$\begin{aligned}\mathbb{E}\left[(\mathcal{E}(L)_\tau)^q\right] &\overset{\text{Hölder}}{\leq} \left(\mathbb{E}\left[\exp\left(\sqrt{qu}L_\tau - \frac{qu}{2}\langle L\rangle_\tau\right)\right]\right)^{\frac{1}{u}} \cdot \left(\mathbb{E}\left[\exp\left(v \cdot \left(q - \sqrt{\frac{q}{u}}\right) L_\tau\right)\right]\right)^{\frac{1}{v}} \\ &\overset{(3.12)}{=} \Bigg(\underbrace{\mathbb{E}\left[\mathcal{E}(\sqrt{qu}L)_\tau\right]}_{(*)}\Bigg)^{\frac{1}{u}} \cdot \Bigg(\underbrace{\mathbb{E}\left[\exp\left(\frac{\sqrt{p}}{2(\sqrt{p}-1)}L_\tau\right]\right)}_{(**)}\Bigg)^{\frac{1}{v}}\end{aligned}$$

As $\mathcal{E}(\sqrt{qu}L)$ is a local martingale that is bounded from below (by zero), it is a supermartingale by Lemma 2.20,

$$(*) \leq \mathbb{E}\left[\mathcal{E}(\sqrt{qu}L)_0\right] = 1.$$

Furthermore, $(**)$ is uniformly bounded for all stopping times $\tau \in \mathcal{T}_T$ by assumption, hence

$$\sup_{\tau \in \mathcal{T}_T} \mathbb{E}\left[(\mathcal{E}(L)_\tau)^q\right] < \infty.$$

By the lemma of de la Vallée Poussin, $\{\mathcal{E}(L)_\tau \mid \tau \in \mathcal{T}_T\}$ is uniformly integrable. As $\mathcal{E}(L)$ is a local martingale, $(\mathcal{E}(L)_{t\wedge\tau_n})_t$ is a martingale for a localizing sequence (τ_n). For the stopped process we therefore have

$$\mathbb{E}\left[\mathcal{E}(L)_{t\wedge\tau_n} \mid \mathcal{F}_s\right] = \mathcal{E}(L)_{s\wedge\tau_n}.$$

The RHS converges to $\mathcal{E}(L)_s$ almost surely as $n \to \infty$. The LHS converges to $\mathbb{E}\left[\mathcal{E}(L)_t \mid \mathcal{F}_s\right]$ in L^1, hence a subsequence converges almost surely. Thus we have verified the martingale property of $\mathcal{E}(L)$, which finalizes the proof. $\square$

Proof of Theorem 3.16.
Let $a \in (0,1)$ and set $p := \frac{1}{(1-a)^2}$ and $q = \frac{p}{p-1}$. Then $a \cdot \frac{\sqrt{p}}{\sqrt{p}-1} = 1$. Thus we have

$$\sup_{\tau \in \mathcal{T}_T} \mathbb{E}\left[\exp\left(\frac{\sqrt{p}}{2(\sqrt{p}-1)} \cdot aL_\tau\right)\right] = \sup_{\tau \in \mathcal{T}_T} \mathbb{E}\left[\exp\left(\frac{1}{2}L_\tau\right)\right] \overset{\text{assumption}}{<} \infty.$$

We are therefore able to apply Lemma 3.17 which tells us that $\mathcal{E}(aL)$ is a uniformly integrable martingale. In order to prove that $M = \mathcal{E}(L)$ is a martingale we calculate how the two are connected:

$$\mathcal{E}(aL) = e^{aL - \frac{a^2}{2}\langle L \rangle} = e^{a^2 L - \frac{a^2}{2}\langle L \rangle} \cdot e^{a(1-a)L} = \mathcal{E}(L)^{a^2} \cdot e^{a(1-a)L}$$

If we take expected values on both sides and apply Hölder's inequality for $\widehat{p} = \frac{1}{a^2}$ and $\widehat{q} = \frac{1}{1-a^2}$ we obtain

$$1 = \mathbb{E}\left[\mathcal{E}(aL)_T\right] \overset{\text{Hölder}}{\leq} \left(\mathbb{E}\left[\mathcal{E}(L)_T\right]\right)^{a^2} \cdot \left(\mathbb{E}\left[e^{\frac{a(1-a)}{1-a^2}L_T}\right]\right)^{1-a^2}$$

With Jensen's inequality[12] we have

$$\mathbb{E}\left[e^{\frac{a(1-a)}{1-a^2}L_T}\right] = \mathbb{E}\left[e^{\frac{a}{1+a}L_T}\right] \leq \left(\mathbb{E}\left[e^{\frac{1}{2}L_T}\right]\right)^{\frac{2a}{1+a}} \overset{\text{assumption}}{<} \infty.$$

Consequently, for any $a \in (0,1)$ we have

$$1 \leq \left(\mathbb{E}\left[\mathcal{E}(L)_T\right]\right)^{a^2} \cdot \left(\mathbb{E}\left[e^{\frac{1}{2}L_T}\right]\right)^{2a(1-a)}.$$

If we let $a \to 1$ we get that $1 \leq \mathbb{E}\left[\mathcal{E}(L)_T\right]$. As $\mathcal{E}(L)$ is a supermartingale (see previous proof) we also have $\mathbb{E}\left[\mathcal{E}(L)_T\right] \leq 1$, hence $M = \mathcal{E}(L)$ satisfies

$$\mathbb{E}\left[M_T\right] = 1 = \mathbb{E}\left[M_0\right],$$

hence M is a martingale on $[0,T]$. $\square$

[12] With $\frac{2a}{1+a} < 1$, $x^{\frac{2a}{1+a}}$ is concave.

Remark 3.18 (Alternative formulation of Kazamaki's condition in the literature)**.** *Let L be a continuous local martingale with $L_0 = 0$. If $(e^{\frac{1}{2}L_t})_t$ is a submartingale, then $\mathcal{E}(L)$ is a martingale.*

For details as to how to prove the statement from the remark, see Chapter VIII, §1, Proposition (1.14) (page 331) in [RY99].
With Remark 3.18 one also has the following result:

Corollary 3.19. *Let L be a continuous martingale with $L_0 = 0$. If*

$$\mathbb{E}\left[e^{\frac{1}{2}L_T}\right] < \infty,$$

then $\mathcal{E}(L)$ is a martingale on $[0, T]$.

Proof of Corollary 3.19. One can show that if X is a submartingale and g a non-decreasing convex function and if $\mathbb{E}[|g(X_t)|] < \infty$ for all t, then $g(X)$ is a submartingale and $\mathbb{E}[g(X_t)] \leq \mathbb{E}[g(X_T)]$ for all $t \leq T$. $g(x) = e^{\frac{1}{2}x}$ is an increasing convex function and with $X = L$ the conditions are satisfied. Therefore,

$$\mathbb{E}\left[e^{\frac{1}{2}L_t}\right] \leq \mathbb{E}\left[e^{\frac{1}{2}L_T}\right] < \infty, \quad \forall t \in [0, T]$$

and thus Remark 3.18 can be applied. Alternatively, the above inequality also holds if t is replaced by a stopping time $\tau \in \mathcal{T}_T$. Thus, Kazamaki's condition is satisfies and the result follows.

$\square$

Theorem 3.20 (Novikov's condition)**.** *Let L be a local martingale with $L_0 = 0$. If*

$$\mathbb{E}\left[e^{\frac{1}{2}\langle L\rangle_T}\right] < \infty,$$

then $M = \mathcal{E}(L)$ is a martingale on $[0, T]$.

Remark 3.21. *If $L_t = \int_0^t \alpha_s \mathrm{d}W_s$ (cf. Theorem 3.15), then Novikov's condition can be rewritten als*

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \alpha_s^2 \mathrm{d}s\right)\right] < \infty.$$

Proof of Novikov's condition. Let $\tau \in \mathcal{T}_T$. We have

$$[\mathcal{E}(L)_\tau]^{\frac{1}{2}} = e^{\frac{1}{2}L_\tau} \cdot \left(e^{-\frac{1}{2}\langle L\rangle_\tau}\right)^{\frac{1}{2}}.$$

As the quadratic variation is non-decreasing, we have

$$e^{\frac{1}{2}L_\tau} = [\mathcal{E}(L)_\tau]^{\frac{1}{2}} \cdot \left[e^{\frac{1}{2}\langle L\rangle_\tau}\right]^{\frac{1}{2}} \leq [\mathcal{E}(L)_\tau]^{\frac{1}{2}} \cdot \left[e^{\frac{1}{2}\langle L\rangle_T}\right]^{\frac{1}{2}}$$

We take expected values on both sides of the inequality. An application of the Cauchy-Schwarz inequality gives

$$\mathbb{E}\left[e^{\frac{1}{2}L_\tau}\right] \leq \mathbb{E}\left[[\mathcal{E}(L)_\tau]^{\frac{1}{2}} \cdot \left[e^{\frac{1}{2}\langle L\rangle_T}\right]^{\frac{1}{2}}\right] \overset{(CSI)}{\leq} \Bigg(\underbrace{\mathbb{E}\left[\mathcal{E}(L)_\tau\right]}_{\leq 1}\Bigg)^{\frac{1}{2}} \Bigg(\underbrace{\mathbb{E}\left[e^{\frac{1}{2}\langle L\rangle_T}\right]}_{<\infty \text{ by ass.}}\Bigg)^{\frac{1}{2}}, \tag{3.13}$$

where we used again that $\mathcal{E}(L)$ is a non-negative (hence bounded from below) local martingale and $\mathbb{E}[\mathcal{E}(L)_0] = 1$; consequently, by Lemma 2.20, it is a supermartingale, hence

$$\mathbb{E}[\mathcal{E}(L)_\tau] = \mathbb{E}[\mathcal{E}(L)_\tau \mid \mathcal{F}_0] \leq \mathcal{E}(L)_0 = 1.$$

On the RHS of (3.13) we have a finite upper bound for any stopping time $\tau \in \mathcal{T}_T$, hence Kazamaki's condition is satisfies. The claim follows with Theorem 3.16. □

Example 3.22. *Let Y satisfy the following SDE for $\mu \in \mathbb{R}$ and $t \in [0, T]$:*

$$\mathrm{d}Y_t = \mu \mathrm{d}t + \mathrm{d}W_t.$$

Then Y is a standard BM w.r.t. the measure $\mathbb{Q}$ defined via

$$\mathrm{d}\mathbb{Q} = \exp\left(-\mu W_T - \frac{1}{2}\mu^2 T\right) \mathrm{d}\mathbb{P}.$$

Theorem 3.23 (Girsanov Theorem II). *Let $Y \in \mathbb{R}^n$ be an Itô process of the form*

$$\mathrm{d}Y_t = \beta_t \mathrm{d}t + \theta_t \mathrm{d}W_t$$

where W is an m-dimensional BM and $\beta_t \in \mathbb{R}^n$, $\theta_t \in \mathbb{R}^{n,m}$ are adapted stochastic processes. Assume that there exist $u \in \mathcal{W}^m$ and $\alpha \in \mathcal{W}^n$ such that

$$\theta_t u_t = \beta_t - \alpha_t.$$

Let

$$M_t := \exp\left(-\sum_{i=1}^m \int_0^t u_s^i \mathrm{d}W_s^i - \frac{1}{2}\int_0^t \|u_s\|^2 \mathrm{d}s\right), \quad t \in [0, T]$$

and define $\mathbb{Q}$ via

$$\mathrm{d}\mathbb{Q} = M_T \mathrm{d}\mathbb{P}.$$

If M is a $\mathbb{P}$-martingale, then $\mathbb{Q}$ is a probability measure and the process

$$\widehat{W}_t := \int_0^t u_s \mathrm{d}s + W_t, \quad t \in [0, T]$$

is a $\mathbb{Q}$-BM. The process Y has the representation

$$\mathrm{d}Y_t = \alpha_t \mathrm{d}t + \theta_t \mathrm{d}\widehat{W}_t.$$

Proof for $n = 1$. We have already seen that a measure $\mathbb{Q}$ defined in this manner is indeed a probability measure. Furthermore, from Girsanov's first theorem (Theorem 3.15) follows that $\widehat{W}$ is a $\mathbb{Q}$-BM — $\widehat{W}$ plays the role of Y in that theorem!
Finally,

$$\begin{aligned}
\mathrm{d}Y_t &= \beta_t \mathrm{d}t + \theta_t \mathrm{d}W_t && \text{replace } \beta \\
&= (\theta_t u_t + \alpha_t)\mathrm{d}t + \theta_t \mathrm{d}W_t && \text{replace } u_t \mathrm{d}t \text{ by } \mathrm{d}\widehat{W}_t - \mathrm{d}W_t \\
&= \theta_t(\mathrm{d}\widehat{W}_t - \mathrm{d}W_t) + \alpha_t \mathrm{d}t + \theta_t \mathrm{d}W_t \\
&= \alpha_t \mathrm{d}t + \theta_t \mathrm{d}\widehat{W}_t.
\end{aligned}$$

□

For details on the multidimensional versions of Girsanov's theorems and their proofs, see e.g. [Øk03], Section 8.6.

Remark 3.24.

(i) If θ is invertible (which requires in particular that $m = n$), then $u_t = \theta_t^{-1}(\beta_t - \alpha_t)$.

(ii) If $\alpha \equiv 0$, then Y has the representation $\mathrm{d}Y_t = \theta_t \mathrm{d}\widehat{W}_t$, hence Y is a local martingale w.r.t. $\mathbb{Q}$. In this case $\mathbb{Q}$ is called an equivalent local martingale measure.

Example 3.25. *Let us once more return to the Black-Scholes model that we have last looked at in Section 2.3.2. Recall that the price process is assumed to be a geometric BM, i.e.,*

$$\mathrm{d}S_t = S_t(\mu \mathrm{d}t + \sigma \mathrm{d}W_t), \quad 0 \leq t \leq T.$$

Furthermore, $S_t^0 = e^{rt}$ for $r > 0$, hence the discounted price process is given by

$$\widetilde{S}_t = e^{-rt} S_t.$$

Introduce an equivalent probability measure $\mathbb{Q} \sim \mathbb{P}$ with density

$$M_T := \exp\left(- \int_0^T \frac{\mu - r}{\sigma} \mathrm{d}W_s - \frac{1}{2} \int_0^T \left(\frac{\mu - r}{\sigma} \right)^2 \mathrm{d}s \right).$$

The quotient $\frac{\mu - r}{\sigma}$ is also known as market price of risk *— it is the difference between the return rates of risky and riskless asset in terms of volatility σ.*
For $\widehat{W}_t = \int_0^t \frac{\mu - r}{\sigma} \mathrm{d}s + W_t$ we have

$$\begin{aligned}
\mathrm{d}\widetilde{S}_t &= e^{-rt} \mathrm{d}S_t - r e^{-rt} S_t \mathrm{d}t \\
&= e^{-rt} \mu S_t \mathrm{d}t + e^{-rt} \sigma S_t \mathrm{d}W_t - r e^{-rt} S_t \mathrm{d}t \\
&= e^{-rt} \left[(\mu - r) S_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t \right] \\
&= \sigma e^{-rt} \left[\frac{\mu - r}{\sigma} S_t \mathrm{d}t + S_t \mathrm{d}W_t \right] \\
&= \sigma e^{-rt} S_t \mathrm{d}\widehat{W}_t \\
&= \sigma \widetilde{S}_t \mathrm{d}\widehat{W}_t.
\end{aligned}$$

Thus, $\widetilde{S} = \mathcal{E}(\sigma \widehat{W})$. Novikov's condition applies, hence $\widetilde{S}$ is a $\mathbb{Q}$-martingale. Therefore, according to the Itô representation theorem (Theorem 2.35), for any $F \in L^2(\mathbb{Q}, \mathcal{F}_T)$, there exists $H \in \mathcal{V}$ such that

$$F = \mathbb{E}^{\mathbb{Q}}[F] + \int_0^T H_t \mathrm{d}\widehat{W}_t = \mathbb{E}^{\mathbb{Q}}[F] + \int_0^T \frac{H_t}{\sigma \widetilde{S}_t} \mathrm{d}\widetilde{S}_t,$$

which is well defined as $\sigma, \widetilde{S}_{\cdot} > 0$. If F is a contingent claim (in the terminology of FiMa I), the process H is a candidate for a strategy replicating F and $\mathbb{E}^{\mathbb{Q}}[F]$ is the arbitrage-free price of F. In the next chapter we will address the problem of arbitrage-free pricing in more detail.

4 NA-Theory and Risk-Neutral Pricing

Throughout this chapter let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space where $(\mathcal{F}_t)$ is generated by an n-dimensional BM W, augmented by the null-sets. The financial market shall consist of $d \leq n$ risky assets with price processes $S^1, \dots, S^d$ and one riskless asset with price process S^0. We assume that $r = 0$.
Before we can work with in this setting, we have to briefly state how we define a multi-dimensional Itô integral.

Definition 4.1. *Let $f\colon [0, \infty) \times \Omega \to \mathbb{R}^{m \times n}$ such that $f_{k,j} \in \mathcal{V}$ for $k \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$. Then we define*

$$\int f \mathrm{d}W := \int \begin{pmatrix} f_{1,1} & \cdots & f_{1,n} \\ \vdots & & \vdots \\ f_{m,1} & \cdots & f_{m,n} \end{pmatrix} \begin{pmatrix} \mathrm{d}W_1 \\ \vdots \\ \mathrm{d}W_n \end{pmatrix}.$$

For $f_{k,j} \in \mathcal{W}$, the integral is defined analogously.
With this, suppose now that the $\mathbb{R}^d$-valued price process S satisfies

$$\mathrm{d}S_t = \gamma_t \mathrm{d}t + \sigma_t \mathrm{d}W_t$$

for $\sigma \in \mathcal{W}^{d \times n}$ and $\gamma \in \mathbb{R}^d$ satisfying $\mathbb{P}\left(\int_0^T |\gamma_s| \mathrm{d}s < \infty\right) = 1$. Hence, S is an $\mathbb{R}^d$-valued Itô process whose k-th component ($k \in \{1, \dots, d\}$) satisfies the 1-dim. SDE

$$\mathrm{d}S_t^k = \gamma_t^k \mathrm{d}t + \sum_{j=1}^n \sigma_t^{k,j} \mathrm{d}W_t^j.$$

Recall that lower indices t and T are used to denote the time, whereas we shall denote by A^T the transpose of a matrix A.

Definition 4.2 (self-financing strategy). *A self-financing strategy (φ^0, φ) is an adapted process which satisfies*

$$\int_0^T |\varphi_t^T \gamma_t| \mathrm{d}t < \infty, \quad \int_0^T \left\|\varphi_t^T \sigma_t\right\|_{\mathbb{R}^n}^2 \mathrm{d}t < \infty \quad \mathbb{P} - a.s.$$

and its associated wealth process $V_t^\varphi = \varphi_t^T S_t + \varphi_t^0$ satisfies $\mathrm{d}V_t^\varphi = \varphi_t^T \mathrm{d}S_t$.

Remark 4.3. *Recall from FiMa I that a strategy $\overline{\xi}$ was self-financing if $\overline{\xi}_t \cdot \overline{S}_t = \overline{\xi}_{t+1} \cdot \overline{S}_t$. If we go from the difference to the differential and adopt our notation, this could be translated as $\mathrm{d}\varphi_t^T \cdot S = 0$. The interpretation is still the same: the change in the value of the portfolio originates from the change in the stock price, not from a change in the strategy.*

Remark 4.4. *We have*

$$V_t^\varphi = \varphi_t^0 + \sum_{k=1}^d \varphi_t^k S_t^k = V_0^\varphi + \sum_{k=1}^d \int_0^t \varphi_s^k \mathrm{d}S_s^k.$$

Hence, for any initial value V_0^φ we can create a self-financing portfolio from $(\varphi^1, \dots, \varphi^d)$ by choosing

$$\varphi_t^0 := V_0^\varphi + \sum_{k=1}^d \left(\int_0^t \varphi_s^k \mathrm{d}S_s^k - \varphi_t^k S_t^k \right).$$

From now on we will therefore restrict our focus on the $\mathbb{R}^d$-valued strategy φ and w.l.o.g. we assume $V_0^\varphi = 0$.

Definition 4.5 (admissible strategy)**.** *An ($\mathbb{R}^d$-valued self-financing) strategy φ is called* admissible *if it is progressively measurable and if*

- *$\varphi^T\sigma \in \mathcal{W}^n$ and $\int_0^T |\varphi_t^T \gamma_t| \mathrm{d}t < \infty$ a.s. and*
- *the associated value process is bounded from below, i.e., there exists $K < \infty$ such that*
$$V_t^\varphi \geq -K \quad \mathbb{P}\text{-a.s. for all } t \in [0, T].$$

The second condition tells us that the trader has a bounded credit limit.

4.1 (No) Arbitrage and existence of ELMM

Definition 4.6. *An admissible strategy φ is called an* arbitrage *if the corresponding value process V^φ satisfies*

$$V_0^\varphi = 0, \quad \mathbb{P}(V_T^\varphi \geq 0) = 1 \text{ and } \mathbb{P}(V_T^\varphi > 0) > 0.$$

Definition 4.7. *A probability measure $\mathbb{Q}$ is called an equivalent local martingale measure (ELMM) if $\mathbb{Q} \approx \mathbb{P}$ and the (discounted) asset price process S is a local martingale under $\mathbb{Q}$.*

Theorem 4.8. *If there exists an ELMM for S, then the model is free of arbitrage.*

Proof. Let $\mathbb{Q} \approx \mathbb{P}$ be an ELMM. Let φ be an admissible trading strategy with associated value process V^φ. As φ is self-financing, by Girsanov V^φ satisfies

$$\mathrm{d}V_t^\varphi = \varphi_t^T \mathrm{d}S_t = \varphi_t^T \sigma_t \mathrm{d}W_t^*$$

for a $\mathbb{Q}$-BM W^*. Then V^φ is a local martingale w.r.t. $\mathbb{Q}$. Since φ is admissible, V^φ is bounded from below, hence it is a supermartingale w.r.t. $\mathbb{Q}$, which implies $0 = V_0^\varphi \geq \mathbb{E}^\mathbb{Q}[V_T^\varphi]$.
Assume that $\mathbb{P}(V_T^\varphi \geq 0) = 1$. Equivalence of $\mathbb{P}$ and $\mathbb{Q}$ implies that also $\mathbb{Q}(V_T^\varphi \geq 0) = 1$, hence we must have $\mathbb{E}^\mathbb{Q}[V_T^\varphi] = 0$ and $\mathbb{Q}(V_T^\varphi = 0) = 1$. Again by equivalence of measures this implies that also $\mathbb{P}(V_T^\varphi = 0) = 1$. This proves NA. □

As in FiMa I, there exists a sufficient condition for the existence of an ELMM, which therefore guarantees NA.

Theorem 4.9. *Assume that there exists a progressively measurable process $\xi = (\xi_t)_{t\in[0,T]}$ such that*

$$\sigma_t(\omega)\xi_t(\omega) = \gamma_t(\omega) \quad \mathbb{P}\otimes\lambda - a.s.,$$

which satisfies the Novikov condition. Then there exists an ELMM.

Proof. Under the assumption on the structure of ξ,

$$\mathrm{d}S_t = \sigma_t(\xi_t \mathrm{d}t + \mathrm{d}W_t) =: \sigma_t \mathrm{d}\widehat{W}_t$$

for $\widehat{W}_t := \int_0^t \xi_s \mathrm{d}s + W_t$. If $M = \mathcal{E}(-\int \xi \mathrm{d}W)$, then (by Girsanov) $\mathrm{d}\mathbb{Q} = M_T \mathrm{d}\mathbb{P}$ defines an equivalent probability measure under which $\widehat{W}$ is a BM.
If $\sigma \in \mathcal{W}$, then S is a $\mathbb{Q}$-local martingale; if the stronger condition $\sigma \in \mathcal{V}$ holds, then S is a $\mathbb{Q}$-martingale. □

As we will shortly see, the converse of the Theorem 4.8 is not true. Before showing this by means of an example let us look at a partial converse statement:

Proposition 4.10. *Suppose that is the market model there is no arbitrage possibility. then there exists an adapted measurable process ξ such that*

$$\sigma\xi = \gamma \quad \mathbb{P}\otimes\lambda\text{-}a.s.$$

Proof. For every $t \in [0, T]$

$$\begin{aligned} F_t :&= \{\omega \in \Omega \,|\, \sigma_t(\omega)\xi_t(\omega) = \gamma_t(\omega) \text{ has no solution}\} \\ &= \{\omega \in \Omega \,|\, \sigma_t\xi_t \neq \gamma_t \text{ for all adapted mb. } \xi\} \\ &= \{\omega \in \Omega \,|\, \gamma_t(\omega) \notin \operatorname{span}\{\sigma_{\cdot,1}, \dots, \sigma_{\cdot,n}\}(\text{columns of } \sigma)\} \\ &= \big\{\omega \in \Omega \,\big|\, \exists v_t(\omega) \text{ s.t. } \sigma_t^T(\omega)v_t(\omega) = 0 \neq v_t^T(\omega)\gamma_t(\omega)\big\} \in \mathcal{F}_t \end{aligned}$$

Let us quickly justify the last equality:

- The image of σ plus the kernel of σ^T equals the entire space $\mathbb{R}^d$. If there exists $\gamma \notin \operatorname{span}\{\dots\}$, γ belongs to the kernel, i.e., $\sigma_t^T\gamma_t = 0$. On the other hand, $\gamma \neq 0$, hence $\gamma^T\gamma > 0$, hence we have $\Rightarrow$ for the choice $v = \gamma$.
- Conversely, let v be such that $\sigma_t^T v_t = 0 \neq v_t^T\gamma_t$ (in ω). Suppose that ξ solves $\sigma\xi = \gamma$. Then $\xi^T\sigma^T = \gamma^T$, hence $\xi^T\sigma^T v = 0 = \gamma^T v \neq 0$, which is a contradiction. Hence we have $\Leftarrow$.

Let $\theta_t^T := \mathbb{1}_{F_t} \operatorname{sgn}(v_t^T\gamma_t)v_t^T$ for some v satisfying $\sigma_t^T v_t = 0 \neq v_t^T\gamma_t$. Then the value process associated to this trading strategy is

$$\begin{aligned} V_T^\theta &= \int_0^T \theta_t^T\gamma_t \mathrm{d}t + \int_0^T \theta_t^T\sigma_t \mathrm{d}W_t \\ &= \int_0^T \mathbb{1}_{F_t} \operatorname{sgn}(v_t^T\gamma_t)v_t^T\gamma_t \mathrm{d}t + \int_0^T \mathbb{1}_{F_t} \operatorname{sgn}(v_t^T\gamma_t)v_t^T\sigma_t \mathrm{d}W_t \\ &= \int_0^T \mathbb{1}_{F_t}|v_t^T\gamma_t|\mathrm{d}t \geq 0 \quad \text{a.s.} \end{aligned}$$

NA implies that $V_T^\theta \equiv 0$ almost surely, hence $(\lambda\otimes\mathbb{P})(F_\cdot) = 0$, which completes the proof. □

Let us proceed with the announced example of an arbitrage-free market in which there exists no ELMM.

Example 4.11 (NA, but no ELMM, source: Delbaen & Schachermayer (1994)). *As usual, let W be a BM and $(\mathcal{F}_t)$ its natural filtration. Let $r = 0$ and fix a time horizon $T = 1$. For $t < 1$ let $f_t := \frac{1}{\sqrt{1-t}}$. With this define*

$$L_t := \exp\left(-\int_0^t f_s \mathrm{d}W_s - \frac{1}{2}\int_0^t f_s^2 \mathrm{d}s\right), \quad 0 \leq t < 1$$

and set $L_1 := \lim_{t\to 1} L_t = 0$. To see that this is true, set $M_t := -\int_0^t f_s \mathrm{d}W_s$. Then

$$\langle M\rangle_t = \int_0^t \frac{1}{1-s}\mathrm{d}s = -\ln(1-t) \xrightarrow{t\to 1} \infty$$

and therefore

$$L_t = e^{M_t - \frac{1}{2}\langle M\rangle_t} = \Bigg(\underbrace{e^{\frac{1}{2}M_t - \frac{1}{8}\langle M\rangle_t}}_{\text{supermart.}}\Bigg)^2 \cdot \underbrace{e^{-\frac{1}{4}\langle M\rangle_t}}_{\to 0 \text{ as } t\to 1}.$$

As a nonnegative stochastic exponential is a nonnegative supermartingale, it converges almost surely and the limit is in L^1. Hence, as the first factor converges and the second factor converges to zero, the product converges to zero as $t \to 1$.
Now define a stopping time $\tau := \inf\{t \geq 0 \mid L_t \geq 2\} \wedge 1$. Then

$$L_\tau = \begin{cases} 2 & \text{if } \tau < 1, \\ 0 & \text{if } \tau = 1. \end{cases}$$

With this we see that the stopped process $(L_t^\tau)_{t\in[0,1]}$ is a bounded martingale. (Why?) Then by the (continuous-time) optional stopping theorem,

$$1 = \mathbb{E}[L_0] = \mathbb{E}[L_\tau] = 2\cdot\mathbb{P}(\tau < 1) + 0\cdot\mathbb{P}(\tau = 1),$$

hence $\mathbb{P}(\tau < 1) = \frac{1}{2}$.
Now consider the price process $S_t := W_{t\wedge\tau} + \int_0^{t\wedge\tau} f_s \mathrm{d}s$ for $t \in [0,1]$. If we denote $\mathcal{G}_t := \mathcal{F}_{t\wedge\tau}$, then S is adapted to $(\mathcal{G}_t)$ with $\mathcal{G}_t \subsetneq \mathcal{F}_t$.

There does not exist an ELMM:
If we let $\mathbb{Q} := L_{1\wedge\tau}\mathbb{P} = L_\tau\mathbb{P}$, then by Girsanov I, S is a standard BM w.r.t. $\mathbb{Q}$, hence in particular it is a martingale. However, as $\mathbb{P}(L_\tau = 0) = \frac{1}{2}$, $\mathbb{Q}$ is not equivalent to $\mathbb{P}$, but only $\mathbb{Q} \ll \mathbb{P}$. From the (later) characterization of all ELMMs we see that, as $\sigma = 1 \in \mathbb{R}$ in this case, there cannot be any other ELMM than the candidate $\mathbb{Q}$.

There is no arbitrage in the market:
Let φ be an admissible strategy and suppose that $\mathbb{P}\left(\int_0^1 \varphi_t \mathrm{d}S_t \geq 0\right) = 1$. Then, by absolute continuity, $\mathbb{Q}\left(\int_0^1 \varphi_t \mathrm{d}S_t \geq 0\right) = 1$. By construction, S is a local martingale w.r.t. $\mathbb{Q}$. Lemma A.1 tells us that the stochastic integral $\int_0^\cdot \varphi_t \mathrm{d}S_t$ is also a local martingale w.r.t. $\mathbb{Q}$. By admissibility of φ it is bounded from below, hence it is a $\mathbb{Q}$-supermartingale. Therefore, $\mathbb{E}^{\mathbb{Q}}\left[\int_0^1 \varphi_t \mathrm{d}S_t\right] \leq 0$, hence $\mathbb{Q}\left(\int_0^1 \varphi_t \mathrm{d}S_t = 0\right) = 1$. This implies in the market equipped with measure $\mathbb{Q}$ there is no arbirage.

We want to show that this also holds under the original measure $\mathbb{P}$. To this end, let $\varepsilon > 0$ and define a stopping time $\nu := \inf\left\{ t \geq 0 \,\middle|\, \int_0^t \varphi_u \mathrm{d}S_u \geq \varepsilon \right\} \wedge 1$. The strategy $\psi_t := \varphi_t \mathbb{1}_{\{t \leq \nu\}}$ for $t \in [0,1]$ is admissible and satisfies

$$\int_0^1 \psi_t \mathrm{d}S_t = \begin{cases} 0 & \textit{if } \nu = 1 \qquad \textit{(because } \int_0^1 \varphi_t \mathrm{d}S_t = 0 \;\; \mathbb{Q}\textit{- a.s.)} \\ \varepsilon & \textit{if } \nu < 1. \end{cases}$$

Under $\mathbb{Q}$ there is no arbitrage, hence $\mathbb{Q}(\nu < 1) = 0$. Consequently,

$$\mathbb{Q}\left(\int_0^t \varphi_u \mathrm{d}S_u \leq \varepsilon \right) = 1 \;\; \forall t < 1.$$

Since we do have $\mathbb{P}|_{\mathcal{G}_t} \approx \mathbb{Q}_{\mathcal{G}_t}$ for $t < 1$, this implies

$$\mathbb{P}\left(\int_0^t \varphi_u \mathrm{d}S_u \leq \varepsilon \right) = 1 \;\; \forall t < 1.$$

Finally, as $t \mapsto \int_0^t \varphi_u \mathrm{d}S_u$ is almost surely continuous, we get $\int_0^1 \varphi_u \mathrm{d}S_u \leq 0$ $\mathbb{P}$-a.s., hence no arbitrage exists under $\mathbb{P}$.

4.2 Complete Markets

In the following we consider a market model with a fixed time horizon T and the price process S. We make the assumption that there exists an ELMM $\mathbb{Q}$.

Definition 4.12.

(i) A (European) contingent claim *is a lower bounded $\mathcal{F}_T$-measurable random variable $H \in L^2(\mathbb{Q})$.*

(ii) A contingent claim H is called attainable *if there exists an admissible strategy φ and a number $z \in \mathbb{R}$ such that*

$$H = z + \int_0^T \varphi_t^T \mathrm{d}S_t \quad \textit{a.s.}$$

In this case φ is called replicating *or* hedging strategy *(or portfolio) for H.*

(iii) The market is called complete *if every contingent claim is attainable.*

Theorem 4.13.
Consider the market model with $\mathrm{d}S_t = \gamma_t \mathrm{d}t + \sigma_t \mathrm{d}W_t$ satisfying the assumptions made above. In particular, assume that there exists an ELMM $\mathbb{Q}$.
The market is complete if and only if σ has a left inverse σ^*, i.e., $\sigma_t^* \sigma_t = I_n$ almost surely. In particular, if $n = d$, then the market is complete if and only if σ_t is invertible $\mathbb{P} \otimes \lambda$-a.s. In any case, there exists a unique replicating strategy for any contingent claim $F \in L^2(\mathbb{Q})$.

Proof. Let $\mathbb{Q}$ be given by $\mathbb{Q} = Z_T \mathbb{P}$ with $Z_t = \mathcal{E}(-\int_0^\cdot \xi_s^T \mathrm{d}W_s)_t$. Then $W_t^* := \int_0^t \xi_s \mathrm{d}s + W_t$ is an (n-dimensional) $\mathbb{Q}$-BM. Under $\mathbb{Q}$ we have $\mathrm{d}S_t = \sigma_t \mathrm{d}W_t^*$. Note that in general we have

$$\mathcal{F}_t^* := \sigma(W_s^*, s \le t) \subsetneq \mathcal{F}_t := \sigma(W_s, s \le t),$$

which prevents us from using a representation theorem right away. Let us now proceed to prove each of the statements of the theorem.

1. Assume that $\sigma^* \sigma = I_n$ and let $H \in L^2(\mathbb{Q})$ be an $\mathcal{F}_T$-measurable contingent claim. Let $Z_t := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t}$, hence in particular

$$\frac{\mathrm{d}\mathbb{P}|\mathcal{F}_t}{\mathrm{d}\mathbb{Q}|\mathcal{F}_t} = \frac{1}{Z_t} = \mathcal{E}(\int_0^\cdot \xi_s^T \mathrm{d}W_s^*)_t.$$

If we define $V_t := \mathbb{E}^{\mathbb{Q}}[H|\mathcal{F}_t]$, then V is a $\mathbb{Q}$-martingale w.r.t. the filtration $(\mathcal{F}_t)$. For $s < t$ we have

$$\mathbb{E}^{\mathbb{P}}[Z_t V_t|\mathcal{F}_s] \overset{\text{Bayes}}{=} \underbrace{\mathbb{E}^{\mathbb{P}}[Z_t|\mathcal{F}_s]}_{=Z_s} \underbrace{\mathbb{E}^{\mathbb{Q}}[V_t|\mathcal{F}_s]}_{=V_s} = Z_s V_s,$$

hence ZV is a $\mathbb{P}$-martingale. From the (multidim.) MRT we infer that there exists $g \in \mathcal{V}^n$ such that

$$Z_t V_t = V_0 + \int_0^t g_s^T \mathrm{d}W_s, \quad t \in [0, T],$$

where we used the fact that $\mathbb{E}^{\mathbb{P}}[Z_0 V_0] = \mathbb{E}^{\mathbb{P}}[Z_0] \mathbb{E}^{\mathbb{Q}}[V_0] = 1 \cdot V_0$. By Itô's product rule we have

$$
\begin{aligned}
\mathrm{d}V_t &= \mathrm{d}(V_t Z_t \cdot \frac{1}{Z_t}) = \frac{1}{Z_t} g_t^T \mathrm{d}W_t + V_t Z_t \cdot \frac{1}{Z_t}\xi_t^T \mathrm{d}W_t^* + \frac{1}{Z_t} g_t^T \xi_t \mathrm{d}t \\
&= \frac{1}{Z_t} g_t^T \underbrace{(\xi_t \mathrm{d}t + \mathrm{d}W_t)}_{=\mathrm{d}W_t^*} + V_t \xi_t^T \mathrm{d}W_t^* \\
&= \left(\frac{1}{Z_t} g_t^T + V_t \xi_t^T\right) \mathrm{d}W_t^* =: \phi_t^T \mathrm{d}W_t^*
\end{aligned}
$$

With the left inverse σ^* of σ let us define $\varphi_t := (\sigma_t^*)^T \phi_t$. Then

$$
\begin{aligned}
V_T^\varphi &= \int_0^T \varphi_t^T \mathrm{d}S_t = \int_0^T \varphi_t^T \sigma_t \mathrm{d}W_t^* = \int_0^T ((\sigma_t^*)^T \phi_t)^T \sigma_t \mathrm{d}W_t^* = \int_0^T (\phi_t)^T \sigma_t^* \sigma_t \mathrm{d}W_t^* \\
&= \int_0^T (\phi_t)^T \mathrm{d}W_t^* = V_T - V_0 = H - \mathbb{E}^{\mathbb{Q}}[H]
\end{aligned}
$$

By construction, $\varphi \in \mathcal{W}^d$ (check!) and $V_t^\varphi = \mathbb{E}^{\mathbb{Q}}[H|\mathcal{F}_t] - \mathbb{E}^{\mathbb{Q}}[H] \geq -2K$ for all $t \in [0, T]$ if $-K$ is the lower bound of H, which exists by assumption for any contingent claim. Thus, φ is an admissible strategy replicating H.

2. Suppose there are two replicating strategies φ and $\widetilde{\varphi}$ for a contingent claim H with associated ϕ and $\widetilde{\phi}$ as before, i.e.,

$$
H = \mathbb{E}^{\mathbb{Q}}[H] + \int_0^T \widetilde{\phi}_t^T \mathrm{d}W_t^* = \mathbb{E}^{\mathbb{Q}}[H] + \int_0^T \phi_t^T \mathrm{d}W_t^*.
$$

Then, by Itô's isometry,

$$
\mathbb{E}^{\mathbb{Q}}\left[\left(\int_0^T (\widetilde{\phi}_t^T - \phi_t^T)\mathrm{d}W_t^*\right)^2\right] = \mathbb{E}^{\mathbb{Q}}\left[\int_0^T \left\|\widetilde{\phi}_t - \phi_t\right\|^2 \mathrm{d}t\right] = 0,
$$

hence $\phi = \widetilde{\phi}$ $\mathbb{Q} \otimes \lambda$-a.s., which implies $\varphi = \widetilde{\varphi}$ $\mathbb{P} \otimes \lambda$-a.s.

3. For the other implication assume that the market is complete. (We want to show that σ_t has a left inverse.) Let $\varphi \in \mathcal{V}^n$ and define $F := \int_0^T \varphi_t^T \mathrm{d}W_t^*$. As φ is square integrable, so is F, i.e., $\mathbb{E}^{\mathbb{Q}}[F^2] < \infty$. F^+ and F^- are contingent claims (because they are nonnegative, hence bounded from below). Therefore, by completeness of the market, there exist admissible replicating strategies ψ^+ and ψ^- such that

$$
\begin{aligned}
F^+ &= \mathbb{E}^{\mathbb{Q}}[F^+] + \int_0^T (\psi_t^+)^T \mathrm{d}S_t = \mathbb{E}^{\mathbb{Q}}[F^+] + \int_0^T (\psi_t^+)^T \sigma_t \mathrm{d}W_t^*, \\
F^- &= \mathbb{E}^{\mathbb{Q}}[F^-] + \int_0^T (\psi_t^-)^T \mathrm{d}S_t = \mathbb{E}^{\mathbb{Q}}[F^-] + \int_0^T (\psi_t^-)^T \sigma_t \mathrm{d}W_t^*.
\end{aligned}
$$

With $\mathbb{E}^{\mathbb{Q}}[F] = 0$ and $\psi := \psi^+ - \psi^-$ we have

$$
F = \int_0^T \varphi_t^T \mathrm{d}W_t^* = \int_0^T \psi_t^T \sigma_t \mathrm{d}W_t^*.
$$

From the uniqueness of the Itô representation of F follows that $\varphi_t^T = \psi_t^T \sigma_t$ $\mathbb{Q} \otimes \lambda$-a.s. As φ was arbitrary, we see that $\operatorname{span} \sigma_t^T = \mathbb{R}^n$ a.s. for all t, hence σ has $\mathbb{P} \otimes \lambda$-a.s. a left inverse.

□

4.2.1 Characterization of ELMMs

We know from FiMa I that equivalent martingale measures are not necessarily unique. The same holds true for ELMMs (unless $n = d$ and σ_t is invertible).
In this section we want to characterize ELMMs by their Radon-Nikodym densities. To get some intuition, let $\mathbb{Q}$ be a fixed ELMM and let φ be an admissible strategy. Then (with $\mathrm{d}S_t = \gamma_t \mathrm{d}t + \sigma_t \mathrm{d}W_t$ as before)

$$\mathrm{d}V_t^\varphi = \varphi_t^T \mathrm{d}S_t = \varphi_t^T \sigma_t \mathrm{d}W_t^* = (\sigma_t^T \varphi_t)^T \mathrm{d}W_t^*$$

for the $\mathbb{Q}$-BM W^*. Consequently, the set of attainable payoffs depends on $C_t := \mathrm{Im}(\sigma_t^T)$. Recall from linear algebra that the direct sum of the image (range) and kernel (nullspace) of a linear transform give the entire co-domain. For $\sigma \in \mathcal{W}^{d\times n}$ this tells us that

$$(\mathrm{Im}(\sigma_t^T))^\perp = \ker(\sigma_t) \quad \text{ and } \quad \mathbb{R}^n = \mathrm{Im}(\sigma_t^T) \oplus \ker(\sigma_t) = C_t \oplus C_t^\perp.$$

If we denote by Π_C the projection on C, then any $z \in \mathbb{R}^n$ has the decomposition[13]

$$z = \Pi_{C_t}(z) + \Pi_{C_t^\perp}(z) = \underbrace{\sigma_t^T(\sigma_t\sigma_t^T)^{-1}\sigma_t z}_{\in \mathrm{Im}(\sigma_t^T)} + \underbrace{(\mathrm{Id} - \sigma_t^T(\sigma_t\sigma_t^T)^{-1}\sigma_t) z}_{\in \ker(\sigma_t)}.$$

Remark 4.14. *If we assume that $\det(\sigma_t\sigma_t^T) > 0$ (i.e. σ_t has full rank), and if we denote by $(\sigma_t^T)^+ := (\sigma_t\sigma_t^T)^{-1}\sigma_t$ the pseudoinverse (or Moore-Penrose inverse) of σ_t^T, then with $\phi_t := \sigma_t^T\varphi_t$ we get the representation*

$$\mathrm{d}V_t^\varphi = \phi_t^T \mathrm{d}W_t^* \quad \textit{and} \quad \varphi_t = (\sigma_t^T)^+ \phi_t.$$

We will now use the decomposition on the integrand defining the Radon-Nikodym derivative.

To this end, let ξ be a progressively measurable process satisfying the Novikov condition and such that $\sigma\xi = \gamma$ $\mathbb{P}\otimes\lambda$-almost surely. Then, by Theorem 4.9, $Z_T := \mathcal{E}(-\int_0^\cdot \xi_t^T \mathrm{d}W_t)_T$ defines an ELMM $\mathbb{P}^*$ and $W_t^* = W_t + \int_0^t \xi_s \mathrm{d}s$ is a $\mathbb{P}^*$-BM.

Moreover, let $\mathbb{Q}$ be an arbitrary ELMM, hence

$$M_t := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t} = \mathcal{E}(+\int_0^\cdot \lambda_s^T \mathrm{d}W_s)_t, \quad t \in [0,T]$$

for some $\lambda \in \mathcal{W}^n$ and $W^\mathbb{Q} := W_t - \int_0^t \lambda_s \mathrm{d}s$ is an n-dimensional $\mathbb{Q}$-BM. Since

$$\mathrm{d}S_t = \sigma_t \mathrm{d}W_t^* = \sigma_t(\xi_t \mathrm{d}t + \mathrm{d}W_t) = \sigma_t(\xi_t \mathrm{d}t + (\mathrm{d}W_t^\mathbb{Q} + \lambda_t \mathrm{d}t)) = \sigma_t((\xi_t + \lambda_t)\mathrm{d}t + \mathrm{d}W_t^\mathbb{Q}),$$

and $\mathbb{Q}$ is by assumption an ELMM, the driver $\sigma_t(\xi_t + \lambda_t)$ must vanish, i.e.,

$$\sigma(\xi + \lambda) = 0 \ \mathbb{P}\otimes\lambda\text{-a.s.}$$

This implies that $\lambda_t = -\xi_t + \eta_t$ with $\eta_t \in \ker(\sigma_t) = C_t^\perp$. Suppose that $\sigma_t\sigma_t^T$ (an $d\times d$-matrix) is invertible. Then, as $\xi \notin \ker(\sigma_t)$ by assumption (namely, $\sigma\xi = \gamma$) and as $C_t \oplus C_t^\perp = \mathbb{R}^n$, we must have $\xi_t \in C_t = \mathrm{Im}(\sigma_t^T)$.

[13] $P := \sigma_t^T(\sigma_t\sigma_t^T)^{-1}\sigma_t$ is a projection from $\mathbb{R}^n$ to $\mathrm{Im}(\sigma_t^T)$. Check that indeed $P^2 = P$ and if $z \in \mathrm{Im}(\sigma_t^T)$, then $Pz = z$. By construction, $Pz \in \mathrm{Im}(\sigma_t^T)$ for any $z \in \mathbb{R}^n$.

Summing up, we have $\xi_t \in C_t$ and $\eta_t \in C_t^\perp$, hence $\xi_t \perp \eta_t$, i.e., $\langle \eta_t, \xi_t \rangle = 0$. With this,

$$\begin{aligned} M_t &= \mathcal{E}\left(\int_0^\cdot (\eta_s - \xi_s)^T \mathrm{d}W_s\right)_t \\ &= \exp\left(\int_0^t (\eta_s - \xi_s)^T \mathrm{d}W_s - \frac{1}{2}\int_0^t \underbrace{\|\eta_s - \xi_s\|^2}_{=\|\eta_s\|^2 - 2\langle \eta_s, \xi_s\rangle + \|\xi_s\|^2} \mathrm{d}s\right) \qquad \Big| \; \langle \eta_s, \xi_s \rangle = 0 \\ &= \mathcal{E}\left(\int_0^\cdot \eta_s^T \mathrm{d}W_s\right)_t \cdot \mathcal{E}\left(-\int_0^\cdot \xi_s^T \mathrm{d}W_s\right)_t \end{aligned}$$

Let us summarize our above findings:

Theorem 4.15. *Assume that there exists a progressively measurable ξ satisfying the Novikov condition such that $\sigma\xi = \gamma$ $\mathbb{P} \otimes \lambda$-a.s.*

(i) Any ELMM $\mathbb{Q}$ has a density process of the form

$$Z_t^{\mathbb{Q}} = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t} = \mathcal{E}\left(\int_0^\cdot \lambda_s^T \mathrm{d}W_s\right)_t = \mathcal{E}\left(\int_0^\cdot \eta_s^T \mathrm{d}W_s\right)_t \cdot \mathcal{E}\left(-\int_0^\cdot \xi_s^T \mathrm{d}W_s\right)_t$$

with $\lambda = -\xi + \eta$, $\eta_t = \Pi_{C_t^\perp}(\lambda_t) \in \ker(\sigma_t)$ and $-\xi_t = \Pi_{C_t}(\lambda_t)$ $\mathbb{P} \otimes \lambda$-a.s. They satisfy $\int_0^T \|\lambda_t\|^2 \,\mathrm{d}t = \int_0^T \|\xi_t\|^2 \,\mathrm{d}t + \int_0^T \|\eta_t\|^2 \,\mathrm{d}t$. In particular, $\eta^T \xi = 0$ and $\eta, \xi \in \mathcal{W}^n$ are unique[14] *$\mathbb{P} \otimes \lambda$-a.s.*

(ii) If $\lambda \in \mathcal{W}^n$ with $-\xi_t = \Pi_{C_t}(\lambda_t)$, $\sigma_t \xi_t = \gamma_t$ and if $Z_t := \mathcal{E}(\int_0^\cdot \lambda_s^T \mathrm{d}W_s)_t$ for $t \in [0, T]$ is a martingale, then $\mathbb{Q} := Z_T \mathbb{P}$ is an ELMM.

Remarks concerning the interpretation:

(i) tells us not only that λ_t has the decomposition into the sum of the projections onto C_t and $C_t^\perp$ — this decomposition is known from linear algebra —, but it also tells us that for a fixed ξ any other ELMM has a density defined for λ whose orthogonal projection is precisely $-\xi$ and not any other element of C_t.

For the interpretation of (ii) recall that in general $\mathcal{E}(\int_0^\cdot \lambda_s^T \mathrm{d}W_s)$ need not define the density of an ELMM. However, if $-\sigma_t \Pi_{C_t}(\lambda_t) = \gamma_t$ and if the stochastic exponential corresponding to λ^T is a martingale, then we do indeed get an ELMM.

[14] Uniqueness follows from $-\xi$ and η being projections of λ onto spaces whose direct sum is $\mathbb{R}^n$.

4.2.2 The range of option prices

Recall from FiMa I that the set of arbitrage free prices corresponds to the expected payoffs under all existing equivalent martingale measures. A similar result holds in our continuous time model.

Let $H \geq 0$ be a European contingent claim in the market with $r = 0$.

- The maximal price a buyer is willing to pay for H is

$$p(H) := \sup\{ y \in \mathbb{R} \mid \exists \varphi \text{ admissible with } V_0^\varphi = 0 \text{ s.t. } V_T^\varphi \geq y - H \text{ a.s.}\}.$$

- The minimal price a seller is willing to accept for H is the minimal superhedging price

$$q(H) := \inf\{ z \in \mathbb{R} \mid \exists \varphi \text{ admissible with } V_0^\varphi = 0 \text{ s.t. } V_T^\varphi \geq H - z \text{ a.s.}\}.$$

Note that if H is attainable, then $p(H) = q(H)$ is the unique arbitrage-free price of H. In general we have

Theorem 4.16. *Let $\mathbb{Q}$ be an ELMM such that $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \mathcal{E}(-\int_0^\cdot \xi_s \mathrm{d}W_s)_T$ for ξ satisfying Novikov's condition and $\sigma\xi = \gamma$. Then*

$$\operatorname{essinf} H \leq p(H) \leq \mathbb{E}^{\mathbb{Q}}[H] \leq q(H) \leq \infty.$$

Reminder: $\operatorname{essinf} H = \sup\{ z \in \mathbb{R} \mid z \leq H \text{ a.s.}\}$.

Proof. By Theorem 4.9, $W_t^{\mathbb{Q}} := \int_0^t \xi_s \mathrm{d}s + W_t$ is a $\mathbb{Q}$-BM and $\mathrm{d}S_t = \sigma_t \mathrm{d}W_t^{\mathbb{Q}}$.
Suppose there exists $y \in \mathbb{R}$ and an admissible strategy φ with $V_0^\varphi = 0$ such that $V_T^\varphi \geq y - H$ almost surely. By definition of the value function this implies that

$$\int_0^T \varphi_u^T \mathrm{d}S_u = \int_0^T \varphi_u^T \sigma_u \mathrm{d}W_u^{\mathbb{Q}} \geq y - H \quad \text{a.s.}$$

By admissibility of φ, the stochastic integral is a local $\mathbb{Q}$-martingale bounded from below, hence it is a supermartingale. Consequently, as the integral starts at 0,

$$\mathbb{E}^{\mathbb{Q}}\left[\int_0^T \varphi_u^T \sigma_u \mathrm{d}W_u^{\mathbb{Q}}\right] \leq 0,$$

hence

$$0 \geq \mathbb{E}^{\mathbb{Q}}\left[\int_0^T \varphi_u^T \mathrm{d}S_u\right] \geq y - \mathbb{E}^{\mathbb{Q}}[H],$$

which implies $\mathbb{E}^{\mathbb{Q}}[H] \geq y$. By taking the supremum over all such $y \in \mathbb{R}$ we get $\mathbb{E}^{\mathbb{Q}}[H] \geq p(H)$ as claimed.

Similarly, if $z + \int_0^T \varphi_u^T \mathrm{d}S_u \geq H$, then $z \geq \mathbb{E}^{\mathbb{Q}}[H]$, which implies $q(H) \geq \mathbb{E}^{\mathbb{Q}}[H]$. If no such pair (z, φ) exists, then $q(H) = \infty$ and there is nothing to show.

Finally, let $\varphi \equiv 0$ and set $y = \operatorname{essinf} H$. Then $y \leq H = H + V_T^\varphi$, which implies $y = \operatorname{essinf} H \leq p(H)$. $\square$

5 Stochastic Optimal Control

In pricing theory, there are risky assets and a riskless bond and the trader's choice is just how much to invest for how long. His decision, however, does not have an influence on the price processes. In contrast, in this chapter we look at problems in which someone's choices have in influence on the processes describing the market. Let us start by looking at some examples of stochastic optimization problems.

For all of the examples we will specify the following details:

How does the state of the system evolve? This is usually given by a stochastic process (giving the state of the system at each time t for each scenario ω) satisfying an SDE, which describes the dynamics of the state process

What is the control and what are the constraints? The state of the system is usually influenced by the control. If a given set of constraints are satisfied, the control will be called admissible.

What is to be optimized? There is a performance / cost criterion, i.e., the control shall be chosen in such a way that e.g. profits shall be maximized or costs minimized.

Useful references for this chapter are chapters 2 and 3 in [Pha09] (main source) and chapter 11 in [Øk03].

5.1 Examples

5.1.1 Portfolio allocation

Consider a financial market with one riskless asset (a savings account) with price process $S^0 > 0$ and n risky assets with associated (n-dimensional) price process S.

An agent can choose how much money he invests in the $n+1$ assets at any given time. If his investment strategy shall be self-financing, then this corresponds to merely decide how much to invest in the risky assets. We will denote by $\alpha_t \in \mathbb{R}^n$ the number of shares of the risky asset(s) at time t.

If an agent possesses α_t shares of the risky assets at time t and if we denote by X_t his wealth at time t, then the number of shares of the riskless asset must be $\frac{X_t - \alpha_t \cdot S_t}{S_t^0}$, provided only self-financing strategies are admitted.

The wealth process evolves according to

$$\mathrm{d}X_t = (X_t - \alpha_t \cdot S_t)\frac{\mathrm{d}S_t^0}{S_t^0} + \alpha_t \cdot \mathrm{d}S_t. \tag{5.1}$$

We have now fixed the state of the system (X), the control (α) and the dynamics of the state depending on the control in Equation (5.1).

The portfolio allocation problem is to choose the *best* investment in the assets, possibly under certain constraints. What is not quite obvious is the definition of *best* in this setting. We will present two different approaches.

Expected Utility Criterion An agent compares random incomes for which he knows only the probability distribution. (This is a big assumption!) Under some conditions on the preferences, the preferences on possible outcomes can be represented by a utility function, which associates higher numbers to preferable outcomes, hence a random income X is preferred to X' if $\mathbb{E}[U(X)] \geq \mathbb{E}[U(X')]$. The expected utility is also called *von-Neumann-Morgenstern* (expected) utility.

The utility function U should be nondecreasing and concave, which corresponds to the intuition that more is better and agents are risk averse.[15] There are different classes of utility function that are frequently used, e.g. power utility

$$U(x) := \begin{cases} \frac{1}{p}(x^p - 1), & x \geq 0, \\ -\infty, & x < 0, \end{cases}$$

for some exponent $p \in (0,1)$ or logarithmic utility

$$U(x) := \ln(x),$$

which corresponds to $p \to 0$ for the power utility function. They are called CRRA utility functions (constant relative risk aversion) because the relative risk aversion $\frac{-xU''(x)}{U'(x)} \equiv 1 - p$ is constant.

In a setting with finite time horizon T, the goal is to maximize $\mathbb{E}\left[U(X_T^\alpha)\right]$ over all admissible controls α if X_T^α is the terminal payoff given control α was chosen:

$$\sup_\alpha \mathbb{E}\left[U(X_T^\alpha)\right]$$

Mean-Variance Criterion If we assume that preferences depend only on expectation (more is better) and variance (less uncertainty is better) of the random terminal position X_T, then the optimization problem becomes

$$\inf_\alpha \left\{ \mathrm{Var}(X_T^\alpha) \mid \mathbb{E}\left[X_T^\alpha\right] = m \right\},$$

i.e., the agent minimizes the variance of the random terminal payoff for a given expected payoff. We can rewrite this as an expected utility function by setting $U(x) = -(\lambda - x)^2$ for some $\lambda \in \mathbb{R}$, since $\mathbb{E}\left[(\lambda - X_T^\alpha)^2\right] = \mathrm{Var}(X_T^\alpha) + (\mathbb{E}\left[X_T^\alpha\right] - \lambda)^2$.

5.1.2 Optimal selling of an asset

Assume that an agent owns an asset with associated price process $X = (X_t)$. The agent wants to sell the asset at the best possible moment, taking into account a fixed transaction fee $K > 0$ and the constant interest rate $\beta > 0$. This problem can be formulated as an optimal stopping problem:

$$\sup_\tau \mathbb{E}\left[e^{-\beta\tau}(X_\tau - K)\right],$$

where the admissible controls could be all stopping times on $[0, \infty]$ or just on a finite interval $[0, T]$ (e.g. the life span of the agent).

[15] If you get 100 Euro and nothing both with probability $\frac{1}{2}$, then the expected income is $\frac{1}{2}(U(100) + U(0))$. If U is concave, then this is lower than $U(50)$. In other words, you prefer a sure payment of 50 Euro over gambling.

5.1.3 Quadratic hedging of options

Consider again risky assets with associated (multi-dim.) price process S and a riskless asset with price process $S^0 > 0$. Let H be a European contingent claim for a maturity $T < \infty$, i.e., H is in particular $\mathcal{F}_T$-measurable. Hedging H means that someone tries to find a self-financing strategy such that the wealth at maturity T coincides with the payoff from H.

Assume that an agent owns α_t shares of the risky assets. If the investment strategy is self-financing, then the value of his portfolio follows the dynamics

$$\mathrm{d}X_t^\alpha = \alpha_t \mathrm{d}S_t + (X_t^\alpha - \alpha_t S_t)\frac{\mathrm{d}S_t^0}{S_t^0}.$$

The goal of the agent is to hedge the claim H in such a manner that the expected value of the quadratic hedging error is minimal, i.e.,

$$\inf_\alpha \mathbb{E}\left[(H - X_T^\alpha)^2\right].$$

5.2 Controlled diffusion processes

As in the last chapter, we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ satisfying the usual assumption where $(\mathcal{F}_t)$ is generated by an n-dimensional BM W. The d-dimensional state process X shall satisfy the SDE

$$\mathrm{d}X_t = \gamma(X_t, \alpha_t)\mathrm{d}t + \sigma(X_t, \alpha_t)\mathrm{d}W_t. \tag{5.2}$$

We make the following assumptions:

Assumption 5.1.

(i) The control $\alpha = (\alpha_t)$ is progressively measurable with values in $A \subset \mathbb{R}^m$.

(ii) The measurable functions $\gamma\colon \mathbb{R}^d \times A \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \times A \to \mathbb{R}^{d\times n}$ satisfy a uniform Lipschitz condition, i.e., there exists $K \geq 0$ such that

$$|\gamma(x,a) - \gamma(y,a)| + |\sigma(x,a) - \sigma(y,a)| \leq K|x-y| \quad \forall x, y, \in \mathbb{R}^d,\ \forall a \in A.$$

As before, we fix a time horizon $T < \infty$ and for $0 \leq t \leq T$ we denote by $\mathcal{T}_{t,T}$ the set of stopping times with values in $[t, T]$.
In addition to the above assumptions we restrict the admissible controls to the set

$$\mathcal{A} := \left\{ \alpha \text{ control process} \,\middle|\, \mathbb{E}\left[\int_0^T |\gamma(0,\alpha_t)|^2 + |\sigma(0,\alpha_t)|^2 \mathrm{d}t\right] < \infty \right\}$$

in order to ensure the existence and uniqueness of a strong solution to SDE (5.2) with initial condition $X_t = x$ for any $(t,x) \in [0,T] \times \mathbb{R}^d$. We denote the (version of) this solution with a.s. continuous paths by $\{\, X_s^{t,x} \mid s \in [t,T]\}$.

Remark 5.2. *In order to have all the assumptions from Theorem 3.3, in particular (ii), i.e.,*

$$\exists C \geq 0 \text{ s.t. } |\gamma(x,\cdot)| + |\sigma(x,\cdot)| \leq C(1+|x|) \quad \forall x, \tag{5.3}$$

observe that if we let $\kappa_t := |\gamma(0,\alpha_t)| + |\sigma(0,\alpha_t)|$ for $\alpha \in \mathcal{A}$, then the above property follows from the Lipschitz property. To be very precise, we should replace Theorem 3.3 by a multi-dimensional one where the coefficients of the SDE may not only depend on time t and the space variable x, but also on ω in order to allow us to add the control. A suitable result can be found in Section 1.3.1 in [Pha09], see in particular Definition 1.3.12 and Theorem 1.3.15 in that book and the remarks made between these two.
In particular, the Theorem 1.3.15 in [Pha09] tells us that we have

$$\mathbb{E}\left[\sup_{s\in[t,T]} |X_s^{t,x}|^2\right] < \infty \tag{5.4}$$

instead of $\mathbb{E}\left[\int_0^T |X_t|^2 \mathrm{d}t\right] < \infty$, which Theorem 3.3 provides.

Now that we have the state process and the control let us introduce the target functional. To this end, let $f\colon [0,T]\times\mathbb{R}^d\times A \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to \mathbb{R}$ be measurable functions describing running and terminal revenues, respectively, such that the gain function can be defined as

$$J(t,x,\alpha) := \mathbb{E}\left[\int_t^T f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + g(X_T^{t,x})\right].$$

In order for the gain function to be well defined we make the following assumptions[16]:

Assumption 5.3. *Assume that one of the following conditions is satisfied:*

(i) *g is lower-bounded* or

(ii) *there exists $C \geq 0$ such that $|g(x)| \leq C(1+|x|^2)$ for all $x \in \mathbb{R}^d$.*

Furthermore, assume that if we denote

$$\mathcal{A}(t,x) := \left\{ \alpha \in \mathcal{A} \,\middle|\, \mathbb{E}\left[\int_t^T |f(s,X_s^{t,x},\alpha_s)|\mathrm{d}s\right] < \infty \right\},$$

then we assume that $\mathcal{A}(t,x) \neq \emptyset$ for all $(t,x) \in [0,T]\times\mathbb{R}^d$.

The objective is to maximize J over all controls $\alpha \in \mathcal{A}(t,x)$. The associated value function is then

$$v(t,x) := \sup\{ J(t,x,\alpha) \mid \alpha \in \mathcal{A}(t,x)\}.$$

Definition 5.4.

(i) *$\widehat{\alpha} \in \mathcal{A}(t,x)$ is called an* optimal control *of $J(t,x,\widehat{\alpha}) = v(t,x)$.*

(ii) *A control α is called a* Markovian control *if it can be represented as $\alpha_s = a(s, X_s^{t,x})$ for some measurable function $a\colon [0,T]\times\mathbb{R}^d \to A$.*

[16] For assumptions on f see Remark B.1.

5.3 The dynamic programming principle (DPP)

Dynamic programming, a principle used in stochastic as well as deterministic optimization. It was developed by Richard Bellman in the late 1950s (see [Bel03]). The idea is that instead of solving the optimization problem on the whole interval $[t, T]$, we proceed as follows:

First, let $\theta \in (t, T)$ we search an optimal control given the state value $X_\theta^{t,x}$, allowing us to compute $v(\theta, X_\theta^{t,x})$. Second, we maximize

$$\mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right]$$

over controls on the subinterval $[t, \theta]$.

For a discrete time setting this would give us a recursive construction of an optimal strategy. In our continuous time setting we will see in the next time its infinitesimal version, the HJB equation. But first, let us state the DPP in our setting and prove why this intuitive principle can be applied here.[17]

Theorem 5.5. *Let* $(t, x) \in [0, T] \times \mathbb{R}^d$. *Then*

$$\begin{aligned} v(t,x) &= \sup_{\alpha\in\mathcal{A}(t,x)} \sup_{\theta\in\mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right] \\ &= \sup_{\alpha\in\mathcal{A}(t,x)} \inf_{\theta\in\mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right]. \end{aligned}$$

Before we prove the theorem, let us remark that the two equalities imply that for any stopping time $\theta \in \mathcal{T}_{t,T}$,

$$v(t,x) = \sup_{\alpha\in\mathcal{A}(t,x)} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right], \tag{5.5}$$

i.e., the theorem tells us that we can indeed split the optimization problem in two parts (on $[t, \theta]$ and on $[\theta, T]$).

Proof. We proceed in two steps:

1. First, we show that

$$v(t,x) \le \sup_{\alpha\in\mathcal{A}(t,x)} \inf_{\theta\in\mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right].$$

To see this, let $\alpha \in \mathcal{A}(t, x)$ be any admissible control. For any $\theta \in \mathcal{T}_{t,T}$ we have

$$X_s^{t,x} = X_s^{\theta, X_\theta^{t,x}} \quad \text{on } \{\theta \le s\}$$

[17] Though the DPP can be applied in many different settings, it is by no means always applicable. If, for instance, the running revenue did not only depend on the current value of the state and control, but on the history of either of these processes, then we would not be able to consider the optimization problem on $[\theta, T]$ as a subproblem of the optimization on $[t, T]$, which allows us to formulate the Bellman equation. For the DPP we therefore require the problem to have what is called an *optimal substructure*.

by pathwise uniqueness of the solution to the SDE. Hence,

$$J(t,x,\alpha) = \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + J(\theta, X_\theta^{t,x}, \alpha)\right] \leq \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right].$$

By taking the corresponding infimum and supremum we therefore get

$$\begin{aligned} J(t,x,\alpha) &\leq \inf_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right] \\ &\leq \sup_{\alpha \in \mathcal{A}(t,x)} \inf_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right]. \end{aligned}$$

By definition of v this directly implies the claim.

2. Now we show that

$$v(t,x) \geq \sup_{\alpha \in \mathcal{A}(t,x)} \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right].$$

Let once again $\alpha \in \mathcal{A}(t,x)$ be any admissible control and $\theta \in \mathcal{T}_{t,T}$. By definition of the value function v, for any $\varepsilon > 0$ (and any $\omega \in \Omega$) there exists $\alpha^\varepsilon \in \mathcal{A}(\theta, X_\theta^{t,x})$ such that

$$J(\theta, X_\theta^{t,x}, \alpha^\varepsilon) \geq v(\theta, X_\theta^{t,x}) - \varepsilon.$$

With this we define (for each $\omega \in \Omega$) a new control

$$\widehat{\alpha}_s := \begin{cases} \alpha_s, & s \in [0, \theta], \\ \alpha_s^\varepsilon, & s \in (\theta, T]. \end{cases}$$

This process can be chosen to be progressively measurable[18], hence $\widehat{\alpha} \in \mathcal{A}(t,x)$. With this control we get

$$\begin{aligned} v(t,x) &\geq J(t,x,\widehat{\alpha}) \\ &= \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + J(\theta, X_\theta^{t,x}, \alpha^\varepsilon)\right] \\ &\geq \mathbb{E}\left[\int_t^\theta f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + v(\theta, X_\theta^{t,x})\right] - \varepsilon. \end{aligned}$$

As $\alpha \in \mathcal{A}(t,x)$ and $\theta \in \mathcal{T}_{t,T}$ were arbitrary we can take the supremum over both sets and thereby the asserted inequality.

□

[18] cf. the measurable selection theorem, e.g. in Chapter 7 in [BS79]

5.4 Hamilton-Jacobi-Bellman equation

The Hamilton-Jacobi-Bellman equation (HJB) is the infinitesimal version of the DPP, i.e., it describes the local behaviour of the value function when we send the stopping time $\theta \in \mathcal{T}_{t,T}$ to t. We will first derive the HJB equation and then verify that the smooth solution to that equation is indeed the value function of the stochastic optimal control problem.

5.4.1 Formal derivation of the HJB equation

Let us consider the (stopping) time $\theta = t + h$ with $h > 0$ and a constant control[19] $\alpha_t \equiv a$ for some arbitrary $a \in A$. Then, by (5.5),

$$v(t,x) \geq \mathbb{E}\left[\int_t^{t+h} f(s, X_s^{t,x}, a)\mathrm{d}s + v(t+h, X_{t+h}^{t,x})\right]. \tag{5.6}$$

By assuming that v is smooth enough, we may apply Itô's formula to get

$$\begin{aligned}
&v(t+h, X_{t+h}^{t,x}) \\
&= v(t,x) + \int_t^{t+h} \frac{\partial v}{\partial t}(s, X_s^{t,x})\mathrm{d}s + \int_t^{t+h} D_x v(s, X_s^{t,x}) \cdot \mathrm{d}X_s^{t,x} + \frac{1}{2}\int_t^{t+h} \operatorname{tr}(D_x^2 v(s, X_s^{t,x})\mathrm{d}\langle X\rangle_s) \\
&= v(t,x) + \int_t^{t+h} \left(\frac{\partial v}{\partial t} + \gamma(x,a)\cdot D_x v + \frac{1}{2}\operatorname{tr}(\sigma(x,a)\sigma^T(x,a)D_x^2 v)\right)(s, X_s^{t,x})\mathrm{d}s \\
&\quad + \int_t^{t+h} D_x v(s, X_s^{t,x})^T \sigma(X_s^{t,x}, a)\mathrm{d}W_s \\
&= v(t,x) + \int_t^{t+h} \left(\frac{\partial v}{\partial t} + \mathcal{L}^a v\right)(s, X_s^{t,x})\mathrm{d}s + \int_t^{t+h} D_x v(s, X_s^{t,x})^T \sigma(X_s^{t,x}, a)\mathrm{d}W_s
\end{aligned}$$

where

$$\mathcal{L}^a v = \gamma(x,a)\cdot D_x v + \frac{1}{2}\operatorname{tr}(\sigma(x,a)\sigma^T(x,a)D_x^2 v)$$

if we denote by $D_x v$ the gradient of v, by $D_x^2 v$ its Hessian and by tr the trace of a matrix.

Remark 5.6. *The operator $\mathcal{L}^a$, which we apply to v, is also known as the* infinitesimal generator *of the stochastic process X.*

By substituting the expression for $v(t+h, X_{t+h}^{t,x})$ into (5.6) we obatin

$$0 \geq \mathbb{E}\left[\int_t^{t+h} \left(f(s, X_s^{t,x}, a) + \left(\frac{\partial v}{\partial t} + \mathcal{L}^a v\right)(s, X_s^{t,x})\right)\mathrm{d}s\right] + \mathbb{E}\left[\int_t^{t+h} D_x v(s, X_s^{t,x})^T \sigma(x,a)\mathrm{d}W_s\right]$$

The stochastic integral is always a local martingale and if it is bounded, then it is even a true martingale, hence its expectation is zero. If we divide both sides by h and let $h \to 0$, then (as $X_t^{t,x} = x$)

$$0 \geq f(t,x,a) + \frac{\partial v}{\partial t}(t,x) + \mathcal{L}^a v(t,x)$$

[19] See Remark B.1 for assumptions on f that guarantee that constant controls are admissible.

Since this holds true for all $a \in A$, we obtain

$$\frac{\partial v}{\partial t}(t,x) + \sup_{a \in A} \left[\mathcal{L}^a v(t,x) + f(t,x,a) \right] \leq 0. \tag{5.7}$$

If $\alpha^* \in \mathcal{A}(t,x)$ was an optimal control, then we would have equality in (5.6), which suggests that the supremum in (5.7) is attained, implying that we should have

$$\frac{\partial v}{\partial t}(t,x) + \sup_{a \in A} \left[\mathcal{L}^a v(t,x) + f(t,x,a) \right] = 0, \qquad \forall (t,x) \in [0,T) \times \mathbb{R}^d. \tag{5.8}$$

If we let $\mathcal{S}_d$ denote the symmetric $d \times d$-matrices, then we define the *Hamiltonian* as a function $H\colon [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{S}_d \to \mathbb{R}$ via

$$H(t,x,p,M) := \sup_{a \in A} \left[\gamma(x,a) \cdot p + \frac{1}{2} \operatorname{tr}(\sigma(x,a)\sigma^T(x,a)M + f(t,x,a) \right].$$

Consequently, (5.8) can be reformulated as

$$\frac{\partial v}{\partial t}(t,x) + H(t,x,D_x v(t,x), D_x^2 v(t,x)) = 0, \qquad \forall (t,x) \in [0,T) \times \mathbb{R}^d, \tag{5.9}$$

which we call the HJB equation. The terminal condition associated to this PDE is (by definition of $v(T,x)$ as supremum of $J(T,x,a)$)

$$v(T,x) = g(x), \quad \forall x \in \mathbb{R}^d.$$

5.4.2 Verification theorem

In the following Theorem we will show that a smooth solution to the HJB equation coincides with the value function and, as a by-product, that the control is Markovian.

Theorem 5.7. *Let $w \in C^{1,2}([0,T) \times \mathbb{R}^d) \cap C^0([0,T] \times \mathbb{R}^d)$ satisfy the quadratic growth condition*

$$|w(t,x)| \leq C(1+|x|^2) \quad \forall (t,x) \in [0,T] \times \mathbb{R}^d.$$

1. *Suppose that*

$$-\frac{\partial w}{\partial t}(t,x) - \sup_{a \in A} \left[\mathcal{L}^a w(t,x) + f(t,x,a) \right] \geq 0, \quad \forall (t,x) \in [0,T) \times \mathbb{R}^d \tag{5.10}$$

$$w(T,x) \geq g(x), \quad \forall x \in \mathbb{R}^d. \tag{5.11}$$

   Then $w \geq v$ on $[0,T] \times \mathbb{R}^d$.

2. *Suppose in addition that $w(T,\cdot) = g$ and that there exists a measurable $\widetilde{\alpha}$ with values in A such that*

   - *$\widetilde{\alpha}$ satisfies the HJB equation, i.e.,*

$$-\frac{\partial w}{\partial t}(t,x) - \sup_{a \in A} \left[\mathcal{L}^a w(t,x) + f(t,x,a) \right] = -\frac{\partial w}{\partial t}(t,x) - \left[\mathcal{L}^{\widetilde{\alpha}(t,x)} w(t,x) + f(t,x,\widetilde{\alpha}(t,x)) \right] = 0;$$

   - *the SDE $\mathrm{d}X_s = \gamma(X_s, \widetilde{\alpha}(s,X_s))\mathrm{d}s + \sigma(X_s, \widetilde{\alpha}(s,X_s))\mathrm{d}W_s$ with initial condition $X_t = x$ admits a unique solution, denoted by $\widetilde{X}^{t,x}$;*
   - *the process $\left\{ \widetilde{\alpha}(s, \widetilde{X}_s^{t,x}) \,\middle|\, s \in [t,T] \right\}$ lies in $\mathcal{A}(t,x)$.*

   Then $w = v$ on $[0,T] \times \mathbb{R}^d$ and $\widetilde{\alpha}$ is an optimal Markovian control.

Remark 5.8.

(i) Before we prove this theorem, let us remark that one can easily apply this theory to a minimization problem, since minimizing J is equivalent to maximizing $-J$. In that case one would simply change $\sup$ *into* $\inf$ *in the definition of the Hamiltonian and the rest remains unchanged.*

(ii) The verification theorem tells us that if the stochastic optimal control problem is solvable and if its value function is smooth enough, then the value function can be characterized as solution of a PDE. However, whether an optimal control exists is a different question not answered here. For the interested reader we refer to a paper by Kushner [Kus75]

Proof.

1. Since $w \in C^{1,2}([0,T) \times \mathbb{R}^d)$, we have for all $(t,x) \in [0,T) \times \mathbb{R}^d$, $\alpha \in \mathcal{A}(t,x)$, $s \in [t,T)$ and any stopping time τ with values in $[t,\infty)$ that, by Itô's formula,

$$w(s \wedge \tau, X^{t,x}_{s\wedge\tau}) = w(t,x) + \int_t^{s\wedge\tau} \frac{\partial w}{\partial t}(u, X^{t,x}_u) + \mathcal{L}^{\alpha_u} w(u, X^{t,x}_u) \mathrm{d}u + \int_t^{s\wedge\tau} D_x w(u, X^{t,x}_u)^T \sigma(X^{t,x}_u, \alpha_u) \mathrm{d}W_u. \tag{5.12}$$

Recall from the definition of $\int_0^T f_t \mathrm{d}W_t$ for $f \in \mathcal{W}$ in Section 2.1.3 that

$$\tau_n := \inf \left\{ t \geq 0 \,\middle|\, \int_0^t (f_s)^2 \mathrm{d}s \geq n \right\} \wedge T$$

is a localizing sequence for f such that $f^n := f \mathbb{1}_{[0,\tau_n]} \in \mathcal{V}$, implying that $(\int_0^\cdot f_t^n \mathrm{d}W_t)$ is a (true) martingale.

With this in mind we define

$$\tau_n := \inf \left\{ s \geq t \,\middle|\, \int_t^s |D_x w(u, X^{t,x}_u)^T \sigma(X^{t,x}_u, \alpha_u)|^2 \mathrm{d}u \geq n \right\} \wedge T.$$

It is a localizing sequence such that for $\tau = \tau_n$, the stochastic integral in (5.12) is a martingale. Its expectation being zero we get

$$\mathbb{E}\left[w(s \wedge \tau_n, X^{t,x}_{s\wedge\tau_n})\right] = w(t,x) + \mathbb{E}\left[\int_t^{s\wedge\tau_n} \frac{\partial w}{\partial t}(u, X^{t,x}_u) + \mathcal{L}^{\alpha_u} w(u, X^{t,x}_u) \mathrm{d}u\right].$$

Recall assumption (5.10), which tells us that

$$\frac{\partial w}{\partial t}(u, X^{t,x}_u) + \mathcal{L}^{\alpha_u} w(u, X^{t,x}_u) + f(u, X^{t,x}_u, \alpha_u) \leq 0, \quad \forall \alpha_u \in \mathcal{A}(t,x).$$

This implies

$$\mathbb{E}\left[w(s \wedge \tau_n, X^{t,x}_{s\wedge\tau_n})\right] \leq w(t,x) - \mathbb{E}\left[\int_t^{s\wedge\tau_n} f(u, X^{t,x}_u, \alpha_u) \mathrm{d}u\right].$$

In order to apply dominated convergence, observe that

$$\left|\int_t^{s\wedge\tau_n} f(u, X_u^{t,x}, \alpha_u)\mathrm{d}u\right| \le \int_t^T \left|f(u, X_u^{t,x}, \alpha_u)\right| \mathrm{d}u \quad \text{and}$$
$$\left|w(s\wedge\tau_n, X_{s\wedge\tau_n}^{t,x})\right| \le C(1 + \sup_{s\in[t,T]} \left|X_s^{t,x}\right|^2),$$

where the RHS of the first inequality is integrable by definition of $\mathcal{A}(t,x)$ and the RHS of the second inequality is integrable because of the quadratic growth condition imposed on w and (5.4). Thus, by dominated convergence, we have the inequality

$$\mathbb{E}\left[w(s, X_s^{t,x})\right] \le w(t,x) - \mathbb{E}\left[\int_t^s f(u, X_u^{t,x}, \alpha_u)\mathrm{d}u\right], \quad \forall \alpha \in \mathcal{A}(t,x). \tag{5.13}$$

By continuity of w and once again by dominated convergence we can let $s \to T$ and get

$$\mathbb{E}\left[g(X_T^{t,x})\right] \le w(t,x) - \mathbb{E}\left[\int_t^T f(u, X_u^{t,x}, \alpha_u)\mathrm{d}u\right], \quad \forall \alpha \in \mathcal{A}(t,x).$$

If we rearrange the summands and take the supremum over all admissible controls $\alpha \in \mathcal{A}(t,x)$, we get that indeed $v \le w$ for all $(t,x) \in [0,T] \times \mathbb{R}^d$.

2. We proceed as in the 1. with $\widetilde{\alpha}$ instead of a general admissible strategy $\alpha \in \mathcal{A}(t,x)$. Then instead of (5.13) we get

$$\mathbb{E}\left[w(s, \widetilde{X}_s^{t,x})\right] = w(t,x) - \mathbb{E}\left[\int_t^s f(u, \widetilde{X}_u^{t,x}, \widetilde{\alpha}(u, \widetilde{X}_u^{t,x}))\mathrm{d}u\right]$$

Once again by letting $s \to T$ we obtain

$$w(t,x) = \mathbb{E}\left[\int_t^T f(u, \widetilde{X}_u^{t,x}, \widetilde{\alpha}(u, \widetilde{X}_u^{t,x}))\mathrm{d}u + g(\widetilde{X}_T^{t,x})\right] = J(t,x,\widetilde{\alpha}) \le v(t,x).$$

With the other inequality from 1. we infer that w coincides with the value function v and that $\widetilde{\alpha}$ is an optimal (Markovian) control.

□

Remark 5.9. *If A is a singleton (and therefore f can be written as independent of a), then the verification theorem is a version of the Feynman-Kac formula, stating (in its most basic form) that the smooth solution to the PDE with terminal condition*

$$\frac{\partial w}{\partial t}(t,x) + \mathcal{L}w(t,x) + f(t,x) = 0, \qquad w(T,x) = g(x)$$

has the representation

$$w(t,x) = \mathbb{E}\left[\int_t^T f(u, X_u^{t,x})\mathrm{d}u + g(X_T^{t,x})\right]$$

for a stochastic process satisfying SDE (5.2). *This result is remarkable as it links parabolic PDEs (which have deterministic solutions) and stochastic processes.*

5.4.3 Application: Optimal portfolio allocation for power utility

Let us recall and refine the model presented in Section 5.1.1. An agent can choose how much to invest in a riskless asset with price process S^0 and a risky asset (only one!) with price process S. We assume that the constant interest rate is $r > 0$ and S is a geometric BM, hence

$$\mathrm{d}S_t^0 = rS_t^0\mathrm{d}t, \qquad \mathrm{d}S_t = \mu S_t\mathrm{d}t + \sigma S_t\mathrm{d}W_t.$$

Recall that if X_t denotes the agent's wealth at time t and α_t the amount invested in the risky asset, then (cf. Equation (5.1))

$$\begin{aligned}\mathrm{d}X_t &= (X_t - \alpha_t S_t)\frac{\mathrm{d}S_t^0}{S_t^0} + \alpha_t\mathrm{d}S_t \\ &= (X_t - \alpha_t S_t)r\mathrm{d}t + \mu\alpha_t S_t\mathrm{d}t + \sigma\alpha_t S_t\mathrm{d}W_t.\end{aligned}$$

If we replace $\alpha_t S_t$ by $\widetilde{\alpha}_t X_t$, then the SDE becomes

$$\mathrm{d}X_t = X_t(\widetilde{\alpha}_t\mu + (1-\widetilde{\alpha}_t)r)\mathrm{d}t + X_t\widetilde{\alpha}_t\sigma\mathrm{d}W_t,$$

where we interpret $\widetilde{\alpha}_t$ as the proportion of the wealth X_t at time t that is invested in the risky asset. This is the structure we assumed in (5.2) with

$$\gamma(x,a) = x(a\mu + (1-a)r) \quad \text{and} \quad \sigma(x,a) = xa\sigma.$$

We assume that the investor faces the constraint that at any time t, the control $\widetilde{\alpha}_t$ must lie in a closed convex subset A of $\mathbb{R}$.[20]
For a given investment strategy $\widetilde{\alpha} \in \mathcal{A}$ we will denote by $X^{t,x}$ the wealth process satisfying the given SDE with initial condition $X_t = x$ (for initial time $t > 0$). The agent wants to maximize the expected utility from the terminal wealth $X_T^{t,x}$. For a given (increasing and concave) utility function U, the value function of the portfolio optimization problem is

$$v(t,x) = \sup_{\widetilde{\alpha}\in\mathcal{A}} \mathbb{E}\left[U(X_T^{t,x})\right], \qquad (t,x) \in [0,T]\times\mathbb{R}_+. \tag{5.14}$$

The associated HJB equation is

$$-\frac{\partial w}{\partial t} - \sup_{a\in A}[\mathcal{L}^a w(t,x)] = 0 \tag{5.15}$$

with terminal condition

$$w(T,x) = U(x), x \in \mathbb{R}_+. \tag{5.16}$$

Let us calculate $\mathcal{L}^a$ in this case:

$$\begin{aligned}\mathcal{L}^a w(t,x) &= \gamma(x,a)\frac{\partial w}{\partial x} + \frac{1}{2}\sigma(x,a)^2\frac{\partial^2 w}{\partial x^2} \\ &= x(a\mu + (1-a)r)\frac{\partial w}{\partial x}(t,x) + \frac{1}{2}x^2a^2\sigma^2\frac{\partial^2 w}{\partial x^2}(t,x)\end{aligned}$$

[20] With the given interpretation as a proportion, one should have $\widetilde{\alpha}_t \in [0,1]$ for any t, so the assumption on A is not very demanding.

Observe that

$$\begin{aligned}\frac{\partial \mathcal{L}^a}{\partial a} w(t,x) &= x(\mu - r)\frac{\partial w}{\partial x}(t,x) + ax^2\sigma^2 \frac{\partial^2 w}{\partial x^2}(t,x) \quad \text{and} \\ \frac{\partial^2 \mathcal{L}^a}{\partial a^2} w(t,x) &= x^2\sigma^2 \frac{\partial^2 w}{\partial x^2}(t,x).\end{aligned}$$

Therefore, $\mathcal{L}^a v(t,x)$ is concave in a, which would permit us to calculate its maximum by differentiation methods, if the value function v is concave in x.

Lemma 5.10. *If the utility function U is increasing and concave on $\mathbb{R}_+$, then so is $x \mapsto v(t,x)$ for any $t \in [0,T]$.*

Proof. Let $\widetilde{\alpha} \in \mathcal{A}$ and $0 \le x \le y$. Write

$$Z_s := X_s^{t,y} - X_s^{t,x}.$$

Then Z satisfies the SDE (with initial condition)

$$\mathrm{d}Z_s = Z_s[(\widetilde{\alpha}_s\mu + (1-\widetilde{\alpha}_s)r)\mathrm{d}s + \widetilde{\alpha}_s\sigma\mathrm{d}W_s], \qquad Z_t = y - x \ge 0.$$

Z is the stochastic exponential of the Itô process with drift $\widetilde{\alpha}_t\mu + (1-\widetilde{\alpha}_t)r$ and volatility $\widetilde{\alpha}_t\sigma$, hence $Z_s \ge 0$ for all $s \ge t$. By definition of Z, this implies that $X_s^{t,y} \ge X_s^{t,x}$ for all $s \ge t$ and by monotonicity of U and of the expectation, we have

$$\mathbb{E}\left[U(X_T^{t,x})\right] \le \mathbb{E}\left[U(X_T^{t,y})\right] \le v(t,y), \qquad \forall\, \widetilde{\alpha} \in \mathcal{A}.$$

By taking the supremum over all admissible strategies $\widetilde{\alpha}$ on the LHS we finally get $v(t,x) \le v(t,y)$.

For the concavity of v, let $x_1, x_2 > 0$ and let $\widetilde{\alpha}^1, \widetilde{\alpha}^2 \in \mathcal{A}$ be two admissible control processes. Fix $\lambda \in [0,1]$ and let $x_\lambda = \lambda x_1 + (1-\lambda)x_2$. Furthermore, for $i \in 1,2$, let X^{t,x_i} denote the wealth process starting from x_i at time t controlled by $\widetilde{\alpha}^i$. With this we define the control

$$\widetilde{\alpha}_s^\lambda := \frac{\lambda X_s^{t,x_1}\widetilde{\alpha}_s^1 + (1-\lambda)X_s^{t,x_2}\widetilde{\alpha}_s^2}{\lambda X_s^{t,x_1} + (1-\lambda)X_s^{t,x_2}},$$

which lies in $\mathcal{A}$ by convexity of A. The process $X^\lambda := \lambda X^{t,x_1} + (1-\lambda)X^{t,x_2}$ satisfies the SDE with initial condition

$$\begin{aligned}\mathrm{d}X_s^\lambda &= X_s^\lambda(\widetilde{\alpha}_s^\lambda\mu + (1-\widetilde{\alpha}_s^\lambda)r)\mathrm{d}s + X_s^\lambda\widetilde{\alpha}_s^\lambda\sigma\mathrm{d}W_s, \quad s \ge t, \\ X_t^\lambda &= x_\lambda.\end{aligned}$$

Hence, X^λ is the wealth process starting from x_λ at time t given the control $\widetilde{\alpha}^\lambda$. As U was assumed to be concave, we have

$$U(\lambda X_T^{t,x_1} + (1-\lambda)X_T^{t,x_2}) \ge \lambda U(X_T^{t,x_1}) + (1-\lambda)U(X_T^{t,x_2}),$$

which implies

$$v(t,x_\lambda) \ge \lambda\mathbb{E}\left[U(X_T^{t,x_1})\right] + (1-\lambda)\mathbb{E}\left[U(X_T^{t,x_2})\right].$$

As $\widetilde{\alpha}^1, \widetilde{\alpha}^2$ were arbitrarily chosen, we see that

$$v(t,\lambda x_1 + (1-\lambda)x_2) \ge \lambda v(t,x_1) + (1-\lambda)v(t,x_2), \quad \forall t.$$

$\square$

Having established that the value function v is indeed concave in x, provided U is concave, let us solve the portfolio optimization problem for power utility. For a given $p \in (0,1)$ let $U(x) = \frac{x^p}{p}$ for $x \geq 0$.[21]

We make the following ansatz: Assume that $w(t,x) = \phi(t)U(x)$ for some positive function ϕ. If we plug in such a w in the HJB equation (5.15), we obtain

$$\begin{aligned} 0 &= \frac{\partial w}{\partial t} + \sup_{a \in A} \left[x(a\mu + (1-a)r) \frac{\partial w}{\partial x}(t,x) + \frac{1}{2} x^2 a^2 \sigma^2 \frac{\partial^2 w}{\partial x^2}(t,x) \right] \\ &= \frac{x^p}{p} \left\{ \phi'(t) + p\phi(t) \sup_{a \in A} \left[a(\mu - r) + r + \frac{1}{2} a^2 \sigma^2 (p-1) \right] \right\}. \end{aligned}$$

If $x \neq 0$ and if we also include the terminal condition (5.16), we obtain the simple ODE

$$0 = \phi'(t) + \eta\phi(t), \quad \phi(T) = 1,$$

where

$$\eta = p \sup_{a \in A} \left[a(\mu - r) + r + \frac{1}{2} a^2 (p-1) \sigma^2 \right].$$

Its solution is $\phi(t) = \exp(\eta(T-t))$, giving us the candidate value function

$$w(t,x) = \exp(\eta(T-t))U(x), \quad (t,x) \in [0,T] \times \mathbb{R}_+.$$

By construction, w is strictly increasing, concave in x and it is a smooth solution of the HJB equation with terminal value. Furthermore, as

$$a \mapsto a(\mu - r) + r + \frac{1}{2} a^2 (p-1) \sigma^2$$

is strictly concave (with second derivative $(p-1)\sigma^2 < 0$), we can calculate its supremum/maximum as

$$\widehat{a} := \frac{\mu - r}{(1-p)\sigma^2},$$

which is simultaneously the maximizer of $\mathcal{L}^a w(t,x)$. The wealth process associated to $\widehat{a}$ solves the SDE

$$\mathrm{d}X_t = X_t(\widehat{a}\mu + (1-\widehat{a})r)\mathrm{d}t + X_t \widehat{a}\sigma \mathrm{d}W_t,$$

hence, by virtue of the verification theorem, w is equal to the value process v. The optimal proportion of wealth invested in the risky asset is constant and equals $\widehat{a}$ and $v(t,x) = w(t,x) = \exp(\eta(T-t))\frac{x^p}{p}$ is also known explicitly with

$$\begin{aligned} \eta &= p \left[\widehat{a}(\mu - r) + r + \frac{1}{2} \widehat{a}^2 (p-1) \sigma^2 \right] \\ &= p \left[\frac{(\mu - r)^2}{(1-p)\sigma^2} - \frac{(\mu - r)^2}{2\sigma^2 (1-p)} + r \right] \\ &= \frac{(\mu - r)^2}{2\sigma^2} \cdot \frac{p}{1-p} + rp. \end{aligned}$$

[21] This is the form of utility used by Merton when he solved this problem in 1969 in *Lifetime Portfolio Selection under Uncertainty: the Continuous-Time Case*.

5.5 Concluding remarks

The topic of optimal control (deterministic or stochastic) could be extended to fill an entire lecture by itself. Therefore, it was necessary to make a choice which aspects to include in a short introduction into this topic. However, there are a few interesting directions from the point of view of stochastic finance that justify to be mentioned. For a broader overview, consult e.g. the survey paper [Kus14] by Harold J. Kushner.

5.5.1 Infinite horizon

The infinite horizon stochastic control problem is that of maximizing

$$J(x, \alpha) := \mathbb{E}\left[\int_0^\infty e^{-\beta s} f(X_s^x, \alpha_s) \mathrm{d}s\right]$$

over all admissible controls α. The following features of this problem should be noted:

- In analogy with the finite horizon setting, the HJB equation requires that the Hamiltonian be maximized over all $a \in A$. We derived the HJB equation by restricting our focus on constant controls. Hence we should require all constant controls $\alpha \equiv a \in A$ to be admissible. For that reason it is necessary to include the discounting factor $e^{-\beta t}$ in the target functional J.
- There is no final payment $g(X_T^{t,x})$ and the running revenues (given by f) are independent of the time t.
- All we looked at so far were finite horizon models and in particular we only gave an existence result (Theorem 3.3) of a solution to an SDE on a finite time interval. For an infinite horizon result, see e.g. Theorem 1.3.15 in [Pha09].

5.5.2 Other solution concepts

Consider the controlled process X satisfying the SDE

$$\mathrm{d}X_s = \alpha_s \mathrm{d}W_s$$

and the stochastic optimal control problem with the value function

$$v(t, x) = \sup_{\alpha \in \mathcal{A}} \mathbb{E}\left[g(X_T^{t,x})\right], \qquad (t, x) \in [0, T] \times \mathbb{R}.$$

Let g^c denote the concave envelope of g, i.e., the smallest concave function above g. Then one can show that the value function should satisfy (cf. Section 3.7 in [Pha09])

$$v(t, x) = g^c(x), \qquad \forall (t, x) \in [0, T) \times \mathbb{R}.$$

However, if $g^c \notin C^2$, then this is not a smooth solution to the HJB equation. Therefore, broader solution concepts can be necessary. Besides the already mentioned weak solution there is another type of solution, the so-called *viscosity solution*. For more information on viscosity solution, see e.g. [CIL92] or Chapter 4 in [Pha09].

5.5.3 Game theory

In stochastic optimal control we usually consider one single agent who chooses his strategy (=control) in a way that optimizes his outcome. If, however, there are several agents who can all individually influence the state process by choosing their strategies, then each one of them can only choose his best strategy *given the choice of the other agents*. If agents decide their strategy without any discussion with other agents, this is a *non-cooperative game*; otherwise it is a cooperative game. The most famous concept for non-cooperative games is a *Nash equilibrium*, which is attained if, provided all other agents stick to their strategies, no single agent has an incentive to deviate from his strategy.

It is beyond the scope of this lecture to introduce the details of game theory. It should simply be mentioned that, if one wants to solve a game, it is necessary to solve each individual's problem given the other agents' strategies, i.e., we have to solve a (stochastic) optimal control problem before we can find an equilibrium for such a game.

5.5.4 Numerical aspects

If a stochastic optimal control problem cannot be solved by hand, one might still be able to obtain a numerical solution. If we are able to numerically solve the HJB equation, then our theory tells us that this solution coincides, under certain conditions, with the value function. Furthermore, the literature providing numerical solutions to stochastic problems is still growing, allowing us to spare the detour of finding a PDE and a suitable solution concept.

5.5.5 Risk minimization

We have used the concept of utility function in our problem of optimal portfolio allocation. A somewhat related concept is that of a *risk measure*. Once a risk measure is fixed, the goal is then to minimize the risk of a future position (that is not known and therefore modelled as a random variable) over all admissible controls. The optimization itself is not different from what we have seen in this chapter, but the notion of a risk measure is interesting enough to justify its thorough introduction. This will be the topic of the next chapter.

6 Risk Measures

We will first introduce static risk measures that assign a number (interpreted as risk) to a (bounded) random variable (interpreted as future position) and then turn to dynamic risk measures that assign to an $\mathcal{F}_T$-measurable random variable a process, which can be interpreted as the risk of the future position at any moment in time. For static risk measures, a good reference is [FS04]. For both static and dynamic risk measures once can consult [Car09].

6.1 Introduction to static risk measures

Let X be a random variable on a space $(\Omega, \mathcal{F})$ (without any explicit probability measure), describing the financial position at the end of a trading period. Let $\mathcal{X}$ denote the linear space of bounded measurable functions from Ω to $\mathbb{R}$. We want to quantify the risk of any given $X \in \mathcal{X}$ by a number $\rho(X)$.

Definition 6.1. *A mapping $\rho\colon \mathcal{X} \to \mathbb{R}$ is called a* monetary risk measure *if it satisfies the following conditions for all $X, Y \in \mathcal{X}$:*

Monotonicity: *If $X \leq Y$, then $\rho(X) \geq \rho(Y)$;*

Cash invariance: *If $m \in \mathbb{R}$, then $\rho(X + m) = \rho(X) - m$.*

The interpretation of cash invariance is that by adding a (non-random) amount of money to a risky position one can reduce the risk of that position by that amount. It implies in particular that for any $X \in \mathcal{X}$ and $m \in \mathbb{R}$ one has

$$\rho(X + \rho(X)) = 0 \quad \text{and} \quad \rho(m) = \rho(0) - m.$$

Remark 6.2. *Monotonicity and cash invariance imply that the any monetary risk measure is Lipschitz continuous w.r.t the supremum norm, i.e.,*

$$|\rho(X) - \rho(Y)| \leq \|X - Y\|, \quad \forall X, Y \in \mathcal{X}.$$

To see this, use that $X \leq Y + \|X - Y\|$, which implies that $\rho(Y) - \|X - Y\| \leq \rho(X)$. Reversing the roles of X and Y gives the desired result.

A monetary risk measure can have other desirable properties, which we define next.

Definition 6.3. *A monetary risk measure $\rho\colon \mathcal{X} \to \mathbb{R}$ is called*

- convex *if $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ for any $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$;*
- normalized *if $\rho(0) = 0$;*
- subadditive *if $\rho(X + Y) \leq \rho(X) + \rho(Y)$ for any $X, Y \in \mathcal{X}$.*

Moreover, a convex monetary risk measure is called coherent *if it is positively homogeneous, i.e., if*

$$\rho(\lambda X) = \lambda\rho(X), \quad \forall X \in \mathcal{X}, \lambda \geq 0.$$

The idea behind convex risk measures is that diversification does not increase the risk.

Observe that positive homogeneity implies that the risk measure is normalized and under positive homogeneity the two properties of convexity and subadditivity are equivalent.

As subadditivity is a (too) strong property, as it requires risks to increase at most linearly in the size of the position, but convexity is a reasonable property for a risk measure to have, we usually do not ask for homogeneity of the risk measures we work with, but merely convexity.

Example 6.4 (entropic risk measure). *Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$. Then entropic risk measure is given by*

$$\rho(X) = \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[e^{-\beta X} \right].$$

The coefficient $\beta > 0$ represents the risk aversion of the agent. We will see later that this is indeed a convex risk measure and we will derive where the name entropic *comes from.*

While the idea behind utility is the quantification of preferences, the idea behind a risk measure is to quantify how much money should be stocked to make a risky position *acceptable* (for instance from the point of view of a supervisor). This motivates the introduction of the set

$$\mathcal{A}_\rho = \{ X \in \mathcal{X} \mid \rho(X) \leq 0 \}$$

of positions which are acceptable under risk measure ρ. We call $\mathcal{A}_\rho$ the *acceptance set* of ρ. The following proposition collects some properties of acceptance sets.

Proposition 6.5. *Let ρ be a monetary risk measure and $\mathcal{A} := \mathcal{A}_\rho$ the associated acceptance set.*

1. *$\mathcal{A} \neq \emptyset$ and it satisfies the following conditions:*
   - *(i)* $\inf \{ m \in \mathbb{R} \mid m \in \mathcal{A} \} > -\infty$*;*
   - *(ii) If $X \in \mathcal{A}$, $Y \in \mathcal{X}$ and $X \leq Y$, then $Y \in \mathcal{A}$;*
   - *(iii) For any $X \in \mathcal{X}$ and $Y \in \mathcal{A}$, the set $\{ \lambda \in [0,1] \mid \lambda X + (1-\lambda) Y \in \mathcal{A} \}$ is closed in* $[0,1]$*. In that sense, $\mathcal{A}$ is closed in $\mathcal{X}$ (w.r.t. the supremum norm).*
2. $\rho(X) = \inf \{ m \in \mathbb{R} \mid m + X \in \mathcal{A} \}$*, i.e., ρ can be recovered from its acceptance set $\mathcal{A}$;*
3. *ρ is a convex monetary risk measure if and only if $\mathcal{A}$ is convex;*
4. *ρ is positively homogeneous if and only if $\mathcal{A}$ is a cone;*
5. *ρ is coherent if and only if $\mathcal{A}$ is a convex cone.*

Reminder: A set $\mathcal{C}$ is a cone if for all $x \in \mathcal{C}$ and $\lambda \geq 0$, one has $\lambda x \in \mathcal{C}$.

Proof.

1. $\mathcal{A} \neq \emptyset$ because $\rho(0) \in \mathcal{A}$. (By transl. invariance, $\rho(0+\rho(0)) = \rho(0) - \rho(0) = 0$.)

   (i) $\inf\{m \in \mathbb{R} \mid m \in \mathcal{A}\} > -\infty$ because for any $\varepsilon > 0$, $\rho(\rho(0) - \varepsilon) = \rho(\rho(0)) + \varepsilon = \varepsilon > 0$, hence $\inf\{m \in \mathbb{R} \mid m \in \mathcal{A}\} > \rho(0) - \varepsilon > -\infty$.

   (ii) If $X \in \mathcal{A}$, then $\rho(X) \leq 0$, hence $\rho(Y) \leq \rho(X) \leq 0$ (by monotonicity), hence $Y \in \mathcal{A}$.

   (iii) Closedness of the set $L := \{\lambda \in [0,1] \mid \lambda X + (1-\lambda)Y \in \mathcal{A}\}$ follows from the Lipschitz continuity of ρ: $\lambda \mapsto \rho(\lambda X + (1-\lambda)Y)$ is continuous, hence the upper contour sets are closed, i.e., in particular L is closed.

2. For any $X \in \mathcal{X}$ we have

$$\begin{aligned}\inf\{m \in \mathbb{R} \mid m + X \in \mathcal{A}\} &= \inf\{m \in \mathbb{R} \mid \rho(m+X) \leq 0\} \\ &= \inf\{m \in \mathbb{R} \mid \rho(X) - m \leq 0\} \\ &= \inf\{m \in \mathbb{R} \mid \rho(X) \leq m\} = \rho(X).\end{aligned}$$

3. If ρ is a convex monetary risk measure and $X, Y \in \mathcal{A}$, then $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y) \leq 0$, hence $\lambda X + (1-\lambda)Y \in \mathcal{A}$ (for any $\lambda \in [0,1]$). On the other hand, assume that $\mathcal{A}$ is convex. Let $X, Y \in \mathcal{A}$, $\lambda \in [0,1]$ and $Z := \lambda X + (1-\lambda)Y$. By assumption, $Z \in \mathcal{A}$. Let $m_X, m_Y \in \mathbb{R}$ such that $\rho(X + m_X) \leq 0$ and $\rho(Y + m_Y) \leq 0$. By cash invariance, $\rho(X) \leq m_X$ and $\rho(Y) \leq m_Y$. Furthermore, we have

$$\rho(Z) - [\lambda m_X + (1-\lambda)m_Y] = \rho(\lambda(X + m_X) + (1-\lambda)(Y + m_Y) \leq 0$$

   (by cash invariance and the convexity of $\mathcal{A}$), hence

$$\rho(Z) \leq \lambda m_X + (1-\lambda)m_Y$$

   for all such m_X and m_Y, hence

$$\begin{aligned}\rho(Z) &\leq \lambda \inf\{m \in \mathbb{R} \mid m + X \in \mathcal{A}\} + (1-\lambda)\inf\{m \in \mathbb{R} \mid m + Y \in \mathcal{A}\} \\ &\overset{2.}{=} \lambda\rho(X) + (1-\lambda)\rho(Y)\end{aligned}$$

   i.e., ρ is convex.

4. ρ is positively homogeneous iff $\rho(\lambda X) = \lambda\rho(X)$ for all $X \in \mathcal{X}$ and $\lambda \geq 0$. This is the case iff $\lambda X \in \mathcal{A} \iff X \in \mathcal{A}$ (for $X \in \mathcal{X}$ arbitrary and for any $\lambda \geq 0$). This is precisely the definition of a cone.

5. ρ is coherent iff (by definition) it is a convex monetary positively homogeneous risk measure. By 3. and 4., this is equivalent to $\mathcal{A}$ being a convex cone.

□

Conversely, for a given set of acceptable positions $\mathcal{A} \subset \mathcal{X}$ we can define $\rho_{\mathcal{A}}(X)$ as the minimal capital requirement that makes X acceptable (in the spirit of 2. from the previous proposition):

$$\rho_{\mathcal{A}}(X) := \inf\{m \in \mathbb{R} \mid m + X \in \mathcal{A}\}. \tag{6.1}$$

Proposition 6.6. *Let $\mathcal{A} \neq \emptyset$ be a subset of $\mathcal{X}$ that satisfies (i) and (ii) from Proposition 6.5. Then $\rho_{\mathcal{A}}$ defined as in* (6.1) *is a monetary risk measure. Furthermore,*

1. *if $\mathcal{A}$ is convex, then $\rho_{\mathcal{A}}$ is a convex risk measure;*

2. *if $\mathcal{A}$ is a cone, then $\rho_{\mathcal{A}}$ is positively homogeneous;*

3. *$\mathcal{A}_{\rho_{\mathcal{A}}} = \mathcal{A}$ if and only if $\mathcal{A}$ is closed w.r.t. $\|\cdot\|$.*

Exercise 11. *Prove Proposition 6.6.*

Example 6.7. *Suppose that we have a fixed probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$. We can say that a position is acceptable if the probability of a loss is bounded by a given level $\lambda \in (0,1)$, i.e., if*

$$\mathbb{P}(X < 0) \leq \lambda.$$

The risk measure corresponding to this acceptance set is called Value at Risk *at level λ and is defined by*

$$V@R_\lambda(X) = \inf\left\{ m \in \mathbb{R} \mid \mathbb{P}(m + X < 0) \leq \lambda \right\}.$$

It is positively homogeneous, but typically not convex. The intuition behind this fact is that the risk measure observes the probability of a loss occurring, but not the actual loss.

Example 6.8. *The* worst case measure *is defined by*

$$\rho_{\max}(X) := -\inf\left\{ X(\omega) \mid \omega \in \Omega \right\},$$

i.e., it is the risk measure associated to the acceptance set of all non-negative functions in $\mathcal{X}$. It is the most conservative among all normalized risk measures as we have $\rho_{\max}(X) \geq \rho(X)$ for any $X \in \mathcal{X}$.
As will be shown in the tutorial, $\rho_{\max}$ is a coherent monetary risk measure and it can be represented as

$$\rho_{\max}(X) = \sup\left\{ \mathbb{E}^{\mathbb{Q}}[-X] \;\middle|\; \mathbb{Q} \in \mathcal{M}_1 \right\},$$

where $\mathcal{M}_1$ is the class of all probability measures on $(\Omega, \mathcal{F})$. This representation is called robust representation *of $\rho_{\max}$.*

Example 6.9. *Let $\mathcal{Q}$ be a set of probability measures on $(\Omega, \mathcal{F})$ and consider a mapping $\gamma \colon \mathcal{Q} \to \mathbb{R}$ with $\sup\left\{ \gamma(\mathbb{Q}) \mid \mathbb{Q} \in \mathcal{Q} \right\} < \infty$. Suppose that a position X is acceptable if*

$$\mathbb{E}^{\mathbb{Q}}[X] \geq \gamma(\mathbb{Q}), \quad \forall \mathbb{Q} \in \mathcal{Q}.$$

The acceptance set $\mathcal{A}$ thus described is convex, hence the corresponding risk measure $\rho = \rho_{\mathcal{A}}$ is convex as well and has the form

$$\rho(X) = \sup\left\{ \gamma(\mathbb{Q}) - \mathbb{E}^{\mathbb{Q}}[X] \;\middle|\; \mathbb{Q} \in \mathcal{Q} \right\}.$$

Alternatively, we can write

$$\rho(X) = \sup\left\{ \mathbb{E}^{\mathbb{Q}}[-X] - \alpha(\mathbb{Q}) \;\middle|\; \mathbb{Q} \in \mathcal{M}_1 \right\}, \tag{6.2}$$

where $\alpha \colon \mathcal{M}_1 \to (-\infty, \infty]$ is defined as

$$\alpha(\mathbb{Q}) := \begin{cases} -\gamma(\mathbb{Q}), & \text{if } \mathbb{Q} \in \mathcal{Q}, \\ +\infty, & \text{else.} \end{cases}$$

α is called a penalty function.

6.2 Robust representation of convex risk measures

The goal of this section is to see whether we have a robust representation as in (6.2) for every convex risk measure.

Let $\mathcal{X}$ denote the set of bounded measurable functions from Ω to $\mathbb{R}$ equipped with the supremum norm $\|\cdot\|$. Furthermore, we denote by $\mathcal{M}_1$ the class of probability measures on $(\Omega, \mathcal{F})$ and by $\mathcal{M}_{1,f}$ the finitely additive set functions $Q\colon \mathcal{F} \to [0,1]$ which are normalized to $Q(\Omega) = 1$. We will use the notation $\mathbb{E}^Q[X] := \int_\Omega X \mathrm{d}Q$.

Remark 6.10. *Let $Q \in \mathcal{M}_{1,f}$ and define*

$$\mathcal{X}_0 := \left\{ \sum_{k=1}^{n} a_k \mathbb{1}_{A_k} \,\middle|\, a_k \in \mathbb{R}, A_k \textit{ disjoint and mb.}, n \in \mathbb{N} \right\} \subset \mathcal{X}.$$

Then for any $F \in \mathcal{X}_0$ one has

$$\mathbb{E}^Q[F] = \sum_{k=1}^{n} a_k Q(A_k) \ (\leq \|F\|_\infty).$$

As $\mathcal{X}_0$ is dense in $\mathcal{X}$, $\mathbb{E}^Q[F]$ can be defined for any $F \in \mathcal{X}$ as the extension of $\mathbb{E}^Q[\cdot]$ from $\mathcal{X}_0$ to $\mathcal{X} = \overline{\mathcal{X}_0}$.

Let $\alpha\colon \mathcal{M}_{1,f} \to \mathbb{R} \cup \{+\infty\}$ be a functional such that $\inf\{\alpha(Q) \mid Q \in \mathcal{M}_{1,f}\} \in \mathbb{R}$.
For each $Q \in \mathcal{M}_{1,f}$, the functional $X \mapsto \mathbb{E}^Q[-X] - \alpha(Q)$ is convex, monotone, and cash invariant on $\mathcal{X}$. These properties are preserved when the supremum is taken over all $Q \in \mathcal{M}_{1,f}$, hence

$$\rho(X) := \sup\left\{ \mathbb{E}^Q[-X] - \alpha(Q) \,\middle|\, Q \in \mathcal{M}_{1,f} \right\} \tag{6.3}$$

defines a convex risk measure on $\mathcal{X}$ such that

$$\rho(0) = -\inf\{\alpha(Q) \mid Q \in \mathcal{M}_{1,f}\}.$$

The functional α will be called *penalty function* for ρ on $\mathcal{M}_{1,f}$ and ρ is said to be *represented by α* on $\mathcal{M}_{1,f}$.

Theorem 6.11. *Any convex risk measure ρ on $\mathcal{X}$ is of the form*

$$\rho(X) = \max\left\{ \mathbb{E}^Q[-X] - \alpha^{\min}(Q) \,\middle|\, Q \in \mathcal{M}_{1,f} \right\}, \qquad X \in \mathcal{X}, \tag{6.4}$$

where the penalty function $\alpha^{\min}$ is given by

$$\alpha^{\min}(Q) := \sup\left\{ \mathbb{E}^Q[-X] \,\middle|\, X \in \mathcal{A}_\rho \right\}, \quad Q \in \mathcal{M}_{1,f}.$$

Moreover, $\alpha^{\min}$ is the minimal penalty function which represents ρ, i.e., any penalty function α for which (6.3) *holds, satisfies*

$$\alpha(Q) \geq \alpha^{\min}(Q), \quad \forall Q \in \mathcal{M}_{1,f}.$$

Proof. Let $X \in \mathcal{X}$. Then $X' := \rho(X) + X \in \mathcal{A}_\rho$, hence

$$\alpha^{\min}(Q) \geq \mathbb{E}^Q[-X'] = \mathbb{E}^Q[-X] - \rho(X), \quad \forall Q \in \mathcal{M}_{1,f}.$$

Rearrangement gives

$$\rho(X) \geq \mathbb{E}^Q[-X] - \alpha^{\min}(Q), \quad \forall Q \in \mathcal{M}_{1,f}.$$

Consequently,

$$\rho(X) \geq \sup\left\{ \mathbb{E}^Q[-X] - \alpha^{\min}(Q) \,\middle|\, Q \in \mathcal{M}_{1,f} \right\}, \quad \forall X \in \mathcal{X}. \tag{6.5}$$

If we can show that

$$\forall X \in \mathcal{X} \ \ \exists Q_X \in \mathcal{M}_{1,f} : \ \rho(X) \leq \mathbb{E}^{Q_X}[-X] - \alpha^{\min}(Q_X), \tag{6.6}$$

then from (6.5) and (6.6) follows

$$\sup_{Q \in \mathcal{M}_{1,f}} \left(\mathbb{E}^Q[-X] - \alpha^{\min}(Q) \right) \leq \rho(X) \leq \mathbb{E}^{Q_X}[-X] - \alpha^{\min}(Q_X)$$

for $Q_X \in \mathcal{M}_{1,f}$, which in turn implies

$$\rho(X) = \max\left\{ \mathbb{E}^Q[-X] - \alpha^{\min}(Q) \,\middle|\, Q \in \mathcal{M}_{1,f} \right\}.$$

Before we prove (6.6), let us verify the minimality of $\alpha^{\min}$:
If α is an arbitrary penalty function for ρ, i.e., for any $X \in \mathcal{X}$ we have in particular

$$\rho(X) \geq \mathbb{E}^Q[-X] - \alpha(Q), \quad \forall Q \in \mathcal{M}_{1,f},$$

hence

$$\alpha(Q) \geq \mathbb{E}^Q[-X] - \rho(X).$$

Taking the supremum over all X, we get

$$\begin{aligned} \alpha(Q) &\geq \sup\left\{ \mathbb{E}^Q[-X] - \rho(X) \,\middle|\, X \in \mathcal{X} \right\} \\ &\geq \sup\left\{ \mathbb{E}^Q[-X] - \rho(X) \,\middle|\, X \in \mathcal{A}_\rho \right\} \\ &\geq \sup\left\{ \mathbb{E}^Q[-X] \,\middle|\, X \in \mathcal{A}_\rho \right\} = \alpha^{\min}(Q). \end{aligned} \tag{6.7}$$

Proof of (6.6)**:**

(i) Fix $X \in \mathcal{X}$. Without loss of generality we may assume that $\rho(X) = 0$. Otherwise, consider $\widetilde{X} := X + \rho(X)$ with $\rho(\widetilde{X}) = 0$ and $\mathbb{E}^Q\left[-\widetilde{X}\right] = \mathbb{E}^Q[-X] - \rho(X)$. With this, (6.6) holds for (X, Q_X) if and only it holds for $(\widetilde{X}, Q_X)$.

(ii) Without loss of generality, $\rho(0) = 0$. Otherwise replace ρ by $\widetilde{\rho}(X) := \rho(X) - \rho(0)$.

(iii) Define $\mathcal{B} := \{Y \in \mathcal{X} \mid \rho(Y) < 0\}$. By assumption, $X \notin \mathcal{B}$ and $\mathcal{B}$ is an open convex (by convexity of ρ) subset of $\mathcal{X}$.

(iv) Recall the **Separating Hyperplanes Theorem**[22]: If E is a Banach space and $\mathcal{B}$, $\mathcal{C}$ are two disjoint convex sets in E, one of which has an interior point, then $\mathcal{B}$ and $\mathcal{C}$ can be separated by a non-zero continuous linear functional ℓ on E, i.e.,

$$\ell(x) \le \ell(y), \quad \forall x \in \mathcal{C},\ \forall y \in \mathcal{B}.$$

(v) By applying the separating hyperplanes theorem to $\mathcal{B}$ and $\mathcal{C} = \{X\}$ we see that there exists a continuous linear functional $\ell \colon \mathcal{X} \to \mathbb{R}$ such that

$$\ell(X) \le \inf \{\ell(Y) \mid Y \in \mathcal{B}\} =: b.$$

(vi) Let $Y \ge 0$ and $\lambda > 0$. Then (by monotonicity and (ii))

$$\rho(1 + \lambda Y) \le \rho(1) = \rho(0) - 1 = -1 < 0,$$

hence $1 + \lambda Y \in \mathcal{B}$. By (v) and by linearity of ℓ,

$$\ell(X) \le \ell(1 + \lambda Y) = \ell(1) + \lambda \ell(Y).$$

If we let $\lambda \to \infty$, then we see that we must have $\ell(Y) \ge 0$.

(vii) As the functional ℓ is not always zero (i.e., $\ell \not\equiv 0$) then there exists $Y \in \mathcal{X}$ such that $\ell(Y) > 0$. With $Y = Y^+ - Y^-$ and step (vi) we see that $\ell(Y^+) > 0$. If we assume (w.l.o.g.) that $\|Y\| \le 1$, then we also have $\ell(1 - Y^+) \ge 0$, hence $\ell(1) = \ell(1 - Y^+) + \ell(Y^+) > 0$.

(viii) By the Riesz representation theorem[23] there exists $Q_X \in \mathcal{M}_{1,f}$ such that

$$\mathbb{E}^{Q_X}[Y] = \frac{\ell(Y)}{\ell(1)}, \quad \forall Y \in \mathcal{X}.$$

(ix) By definition of $\mathcal{B}$ (in (iii)), all elements of $\mathcal{B}$ are acceptable, i.e., $\mathcal{B} \subset \mathcal{A}_\rho$, hence

$$\begin{aligned} \alpha^{\min}(Q_X) &= \sup \left\{ \mathbb{E}^{Q_X}[-Y] \;\middle|\; Y \in \mathcal{A}_\rho \right\} \\ &\ge \sup \left\{ \mathbb{E}^{Q_X}[-Y] \;\middle|\; Y \in \mathcal{B} \right\} \\ &= -\inf \left\{ \frac{\ell(Y)}{\ell(1)} \;\middle|\; Y \in \mathcal{B} \right\} \overset{(v)}{=} -\frac{b}{\ell(1)}. \end{aligned}$$

(x) If $Y \in \mathcal{A}_\rho$, then $Y + \varepsilon \in \mathcal{B}$ for all $\varepsilon > 0$, hence we have equality in (ix), i.e.,

$$\alpha^{\min}(Q_X) = -\frac{b}{\ell(1)}.$$

With this equality and (viii) we get

$$\mathbb{E}^{Q_X}[-X] - \alpha^{\min}(Q_X) = \frac{1}{\ell(1)}(b - \ell(X)) \overset{(v)}{\ge} 0 = \rho(X),$$

which completes the proof of (6.6).

[22] This is Theorem A.58 in the Appendix of [FS04] or Theorem V.2.8 in *Linear Operators. Part I: General Theory* (1958) by N. Dunford and J. Schwartz.

[23] cf. for instance Theorem A.54 in [FS04]

□

Remark 6.12.

1. *If we take $\alpha = \alpha^{\min}$ in (6.7), then all inequalities turn into identities. Thus, we obtain an alternative formula for the minimal penalty function:*

$$\alpha^{\min}(Q) = \sup\left\{\mathbb{E}^Q[-X] - \rho(X) \,\middle|\, X \in \mathcal{X}\right\}.$$

Thus, $\alpha^{\min}$ can be identified as the convex conjugate (or Fenchel-Legendre transform) of ρ.

2. *Suppose that $\rho := \rho_{\mathcal{A}}$ is defined via a given acceptance set $\mathcal{A} \subset \mathcal{X}$ (cf. (6.1)). Then $\mathcal{A}$ directly determines $\alpha^{\min}$:*

$$\alpha^{\min}(Q) = \sup\left\{\mathbb{E}^Q[-X] \,\middle|\, X \in \mathcal{A}\right\}, \quad \forall Q \in \mathcal{M}_{1,f}.$$

This is true because $X \in \mathcal{A}_\rho$ implies that $X + \varepsilon \in \mathcal{A}$ for all $\varepsilon > 0$.

Corollary 6.13. *The minimal penalty function $\alpha^{\min}$ of a coherent risk measure ρ takes only values in $\{0, +\infty\}$. In particular,*

$$\rho(X) = \max\left\{\mathbb{E}^Q[-X] \,\middle|\, Q \in \mathcal{Q}^{\max}\right\}, \quad X \in \mathcal{X}, \tag{6.8}$$

for the convex set

$$\mathcal{Q}^{\max} := \left\{Q \in \mathcal{M}_{1,f} \,\middle|\, \alpha^{\min}(Q) = 0\right\}$$

and $\mathcal{Q}^{\max}$ is the largest set for which (6.8) holds.

Proof. Recall that the acceptance set of a coherent risk measure is a cone. Thus, the minimal penalty function satisfies

$$\alpha^{\min}(Q) = \sup_{X \in \mathcal{A}_\rho} \mathbb{E}^Q[-X] = \sup_{\lambda X \in \mathcal{A}_\rho} \mathbb{E}^Q[-\lambda X] = \lambda \alpha^{\min}(Q)$$

for all $Q \in \mathcal{M}_{1,f}$ and $\lambda > 0$. Hence, $\alpha^{\min}(Q) \in \{0, +\infty\}$. Consequently, as every coherent risk measure is convex, from (6.8) and (6.4) follows that only those $Q \in \mathcal{M}_{1,f}$ should be considered in the robust representation that satisfy $\alpha^{\min}(Q) = 0$. □

Remark 6.14. *The converse of Corollary 6.13 also holds, i.e., if $\mathcal{Q} \subset \mathcal{M}_{1,f}$ and $\rho(X) = \sup\left\{\mathbb{E}^Q[-X] \,\middle|\, Q \in \mathcal{Q}\right\}$ for all $X \in \mathcal{X}$, then ρ is a coherent risk measure.*

As promised, we want to come back to our first example of a risk measure, the *entropic risk measure* (Example 6.4). Let us first introduce the relative entropy of to probability measures, which, broadly speaking, measures how one probability measure diverges from another one.

Definition 6.15. *The relative entropy of a probability measure* $\mathbb{Q}$ *with respect to a given probability measure* $\mathbb{P}$ *is defined as*

$$H(\mathbb{Q}|\mathbb{P}) := \begin{cases} \mathbb{E}^{\mathbb{P}}\left[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \ln\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right)\right], & \text{if } \mathbb{Q} \ll \mathbb{P}, \\ +\infty, & \text{else.} \end{cases}$$

Remark 6.16. *If* $\mathbb{Q} \ll \mathbb{P}$*, then the relative entropy can be rewritten as*

$$H(\mathbb{Q}|\mathbb{P}) = \mathbb{E}^{\mathbb{Q}}\left[\ln\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right)\right].$$

Lemma 6.17. *The relative entropy has the following properties:*

1. $H(\mathbb{Q}|\mathbb{P}) \geq 0$ *for all* $\mathbb{Q} \in \mathcal{M}_1$*;*

2. *For any probability measure* $\mathbb{Q}$*,*

$$\begin{aligned} H(\mathbb{Q}|\mathbb{P}) &= \sup\left\{\mathbb{E}^{\mathbb{Q}}[Z] - \ln \mathbb{E}^{\mathbb{P}}\left[e^Z\right] \;\middle|\; Z \in L^\infty(\mathbb{P})\right\} \\ &= \sup\left\{\mathbb{E}^{\mathbb{Q}}[Z] - \ln \mathbb{E}^{\mathbb{P}}\left[e^Z\right] \;\middle|\; e^Z \in L^1(\mathbb{P})\right\}. \end{aligned}$$

   The second supremum is attained by $Z := \ln \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$ *if* $\mathbb{Q} \ll \mathbb{P}$*.*

These statements can be found as Remark 3.21 and Lemma 3.29 in [FS04].

Proof.

1. The function $\mathbb{R}_{\geq 0} \ni x \mapsto h(x) := x \ln x$ (with $h(0) = 0$) is strictly convex (with second derivative $h''(x) = \frac{1}{x} > 0$ for any $x > 0$). Hence, by Jensen's inequality we have for $\mathbb{Q} \ll \mathbb{P}$

$$H(\mathbb{Q}|\mathbb{P}) = \mathbb{E}^{\mathbb{P}}\left[h\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right)\right] \geq h\left(\mathbb{E}^{\mathbb{P}}\left[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right]\right) = h(1) = 0$$

   with equality if and only if $\mathbb{Q} = \mathbb{P}$.

2. "$\geq$": If $H(\mathbb{Q}|\mathbb{P}) = \infty$, then there is nothing to show. Hence, we assume that $H(\mathbb{Q}|\mathbb{P}) < \infty$. Let Z be such that $e^Z \in L^1(\mathbb{P})$ and define $\mathbb{P}^Z$ via

$$\frac{\mathrm{d}\mathbb{P}^Z}{\mathrm{d}\mathbb{P}} = \frac{e^Z}{\mathbb{E}^{\mathbb{P}}\left[e^Z\right]}.$$

   Then $\mathbb{P}^Z$ is equivalent to $\mathbb{P}$ and

$$\ln \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \ln \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}^Z} + \ln \frac{\mathrm{d}\mathbb{P}^Z}{\mathrm{d}\mathbb{P}}.$$

   Taking $\mathbb{Q}$-expectations on both sides gives

$$H(\mathbb{Q}|\mathbb{P}) = H(\mathbb{Q}|\mathbb{P}^Z) + \mathbb{E}^{\mathbb{Q}}[Z] - \ln \mathbb{E}^{\mathbb{P}}\left[e^Z\right].$$

Since $H(\mathbb{Q}|\mathbb{P}^Z) \geq 0$ by 1., we get that

$$H(\mathbb{Q}|\mathbb{P}) \geq \mathbb{E}^{\mathbb{Q}}[Z] - \ln \mathbb{E}^{\mathbb{P}}\left[e^Z\right], \quad \forall e^Z \in L^1(\mathbb{P}),$$

which implies

$$H(\mathbb{Q}|\mathbb{P}) \geq \sup\left\{\mathbb{E}^{\mathbb{Q}}[Z] - \ln \mathbb{E}^{\mathbb{P}}\left[e^Z\right] \;\middle|\; e^Z \in L^1(\mathbb{P})\right\}.$$

"$\leq$": If $\mathbb{Q} \not\ll \mathbb{P}$, then there exists A such that $\mathbb{Q}(A) > 0$ and $\mathbb{P}(A) = 0$. For $Z_n := n\mathbb{1}_A$ we have that

$$\mathbb{E}^{\mathbb{Q}}[Z_n] - \ln \mathbb{E}^{\mathbb{P}}\left[e^{Z_n}\right] = n \cdot \mathbb{Q}(A) \overset{n\to\infty}{\longrightarrow} H(\mathbb{Q}|\mathbb{P}),$$

which proves the claim.

Now assume that $\mathbb{Q} \ll \mathbb{P}$ with density $\varphi := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$. Then $Z := \ln\varphi$ satisfies $e^Z = \varphi \in L^1(\mathbb{P})$ and we have

$$H(\mathbb{Q}|\mathbb{P}) = \mathbb{E}^{\mathbb{Q}}[Z] - \ln \mathbb{E}^{\mathbb{P}}\left[e^Z\right],$$

which proves the extra statement about where the second supremum is attained, provided we have " $\leq$ " for all other Z.

To see that we have " $\leq$ " in the first line, consider

$$Z_n := \min\{n, \max\{-n, \ln\varphi\}\} \in L^\infty(\mathbb{P}).$$

We claim that

$$\mathbb{E}^{\mathbb{P}}\left[e^{Z_n}\right] \overset{n\to\infty}{\longrightarrow} \mathbb{E}^{\mathbb{P}}\left[e^{\ln\varphi}\right] = 1. \tag{6.9}$$

To see why this holds, split the expectation as follows:

$$\mathbb{E}^{\mathbb{P}}\left[e^{Z_n}\right] = \mathbb{E}^{\mathbb{P}}\left[e^{Z_n}\mathbb{1}_{\{\varphi\geq 1\}}\right] + \mathbb{E}^{\mathbb{P}}\left[e^{Z_n}\mathbb{1}_{\{\varphi<1\}}\right].$$

The first converges by monotone converges, the second by dominated convergence. Now observe that, since $h(x) = x\ln x \geq -\frac{1}{e}$ for every $x \geq 0$, we have $\varphi Z_n \geq -\frac{1}{e}$ for every $n \in \mathbb{N}$. Hence Fatou's lemma (for a sequence of random variables that is uniformly bounded from below) yields

$$\liminf_{n\to\infty} \mathbb{E}^{\mathbb{Q}}[Z_n] = \liminf_{n\to\infty} \mathbb{E}^{\mathbb{P}}[\varphi Z_n] \overset{\text{Fatou}}{\geq} \mathbb{E}^{\mathbb{P}}[\varphi\ln\varphi] = H(\mathbb{Q}|\mathbb{P}). \tag{6.10}$$

By combining (6.9) and (6.10) we see that

$$\liminf_{n\to\infty}\left(\mathbb{E}^{\mathbb{Q}}[Z_n] - \ln \mathbb{E}^{\mathbb{P}}\left[e^{Z_n}\right]\right) \geq H(\mathbb{Q}|\mathbb{P}),$$

which finishes the proof of " $\leq$ ".

□

Example 6.18. *Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$. For a constant $\beta > 0$ define a penalty function $\alpha \colon \mathcal{M}_1 \to \mathbb{R}$ by*

$$\alpha(\mathbb{Q}) := \frac{1}{\beta} H(\mathbb{Q}|\mathbb{P}).$$

The entropic risk measure is given by

$$\rho(X) := \sup \left\{ \mathbb{E}^{\mathbb{Q}}\left[-X\right] - \frac{1}{\beta} H(\mathbb{Q}|\mathbb{P}) \;\middle|\; \mathbb{Q} \in \mathcal{M}_1 \right\}$$

for any $X \in \mathcal{X}$. From part 2 of the above lemma, we know that for any $Z \in L^\infty(\mathbb{P})$,

$$H(\mathbb{Q}|\mathbb{P}) \geq \mathbb{E}^{\mathbb{Q}}\left[Z\right] - \ln \mathbb{E}^{\mathbb{P}}\left[e^{Z}\right],$$

hence in particular for $Z = -\beta X$,

$$H(\mathbb{Q}|\mathbb{P}) \geq \mathbb{E}^{\mathbb{Q}}\left[-\beta X\right] - \ln \mathbb{E}^{\mathbb{P}}\left[e^{-\beta X}\right].$$

Division by β and rearrangement gives

$$\mathbb{E}^{\mathbb{Q}}\left[-X\right] - \frac{1}{\beta} H(\mathbb{Q}|\mathbb{P}) \leq \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}}\left[e^{-\beta X}\right].$$

Part 2 of the above lemma also states that we have equality for $Z = -\beta X = \ln \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}$, which holds if $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = e^{-\beta X}$.
Consequently, ρ has the representation introduced in Example 6.4, i.e.,

$$\rho(X) = \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}}\left[e^{-\beta X}\right], \quad X \in \mathcal{X}.$$

The function $\alpha = \frac{1}{\beta} H(\mathbb{Q}|\mathbb{P})$ is even the minimal penalty function representing ρ, since

$$\alpha^{\min}(\mathbb{Q}) = \sup_X \big(\mathbb{E}^{\mathbb{Q}}\left[-X\right] - \rho(X)\big) = \sup_X \left(\mathbb{E}^{\mathbb{Q}}\left[-X\right] - \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}}\left[e^{-\beta X}\right]\right) = \frac{1}{\beta} H(\mathbb{Q}|\mathbb{P}).$$

6.3 Static and Dynamic Risk Measures for Processes

For this section we follow [PR15]. A good source for the link to BSDEs, which we will also come across, is [Car09] (Chapter 3 and in particular Section 3.6).

Let us fix the setting for this section:
Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a filtered probability space satisfying the usual assumptions with time horizon $T \in [0, \infty]$. By $\mathcal{R}^\infty$ we shall denote the set of all adapted càdlàg processes X that are essentially bounded[24], i.e.,

$$\|X\|_{\mathcal{R}^\infty} := \|X^*\|_{L^\infty} < \infty, \qquad \text{where} \quad X^* := \sup\{|X_t| \mid t \in [0,T]\}.$$

Definition 6.19. *A map $\rho\colon \mathcal{R}^\infty \to \mathbb{R}$ is called a* monetary convex risk measure for processes *if it satisfies the following properties for all $X, Y \in \mathcal{R}^\infty$:*

Cash invariance: $\rho(X + m\mathbb{1}_{[0,T]}) = \rho(X) - m$ *for all* $m \in \mathbb{R}$;

Monotonicity: $\rho(X) \geq \rho(Y)$ *if* $X \leq Y$;

Convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ *for all* $\lambda \in [0,1]$;

Normalization: $\rho(0) = 0$.

A convex monetary risk measure is called coherent*, if in addition to the previous properties it is positive homogeneous, i.e., for all $\lambda \geq 0$,*

$$\rho(\lambda X) = \lambda\rho(X).$$

As we have seen in the last section, we can discard of normalization, but it makes computations easier. It is just for convenience that we require normalization.

Remark 6.20. *As before, we can define the* acceptance set *of a risk measure ρ as*

$$\mathcal{A} := \{X \in \mathcal{R}^\infty \mid \rho(X) \leq 0\}.$$

Using the cash invariance, we get the representation

$$\rho(X) = \inf\left\{m \in \mathbb{R} \,\middle|\, X + m\mathbb{1}_{[0,T]} \in \mathcal{A}\right\}.$$

In other words, $\rho(X)$ is the minimal capital requirement that has to be added to X at time 0 *in order to make it acceptable.*

The main difference between the risk measure as we introduced it for random variables and the (still static) risk measure for stochastic process is that the timing of the cash flow matters. In other words, this new notion of a risk measure also captures the *time value of money*.

To see this more clearly, let us introduce some other concepts of what can happen if money is added at a certain moment in time.

[24] Observe that X^* is defined ω-wise, hence there is not essential supremum. The essential boundedness comes from the L^∞-norm!

Definition 6.21. *A convex risk measure ρ for processes in $\mathcal{R}^\infty$ is called*

- cash subadditive*, if for all $t \geq 0$ and $m \in \mathbb{R}$,*

$$\rho(X + m\mathbb{1}_{[t,T]}) \geq \rho(X) - m, \quad \text{if } m \geq 0,$$
$$\rho(X + m\mathbb{1}_{[t,T]}) \leq \rho(X) - m, \quad \text{if } m \leq 0.$$

- cash additive at t *for some $t > 0$, if*

$$\rho(X + m\mathbb{1}_{[t,T]}) = \rho(X) - m, \qquad \forall m \in \mathbb{R};$$

- cash additive*, if it is cash additive at all times $t \in [0, T]$.*

From monotonicity and cash invariance we can directly infer the following result:

Proposition 6.22. *Every convex (and normalized) monetary risk measure for processes is cash subadditive.*

In contrast, a static risk measure of random variables $\widetilde{\rho}$, if we interpret it as risk measure of processes via $\rho(X) := \widetilde{\rho}(X_T)$, is cash additive.

6.3.1 Dynamic risk measures

In contrast to the static risk measure introduced above, a dynamic risk measure includes a risk assessment at any time $t \in [0, T]$, which shall take into account the information available up to that time.

For a fixed time horizon $T < \infty$ and $0 \leq t \leq s \leq T$ we define the projections

$$\pi_{t,s} \colon \mathcal{R}^\infty \to \mathcal{R}^\infty, \quad \text{with} \quad \pi_{t,s}(X)_r := \mathbb{1}_{[t,T]}(r) X_{r \wedge s}, \quad r \in [0, T].$$

We use the notations

$$\mathcal{R}^\infty_{t,s} := \pi_{t,s}(\mathcal{R}^\infty) \quad \text{and} \quad \mathcal{R}^\infty_t := \pi_{t,T}(\mathcal{R}^\infty).$$

Definition 6.23. *A map $\rho_t \colon \mathcal{R}^\infty_t \to L^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$ for $t \in (0, T]$ is called a* conditional convex risk measure for processes *if it satisfies the following properties for all $X, Y \in \mathcal{R}^\infty_t$:*

conditional cash invariance: *for all $m \in L^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$,*

$$\rho_t(X + m\mathbb{1}_{[t,T]}) = \rho_t(X) - m;$$

monotonicity: *$\rho_t(X) \geq \rho_t(Y)$ if $X \leq Y$;*

conditional convexity: *for all $\lambda \in L^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$ with $\lambda \in [0, 1]$*

$$\rho_t\left(\lambda X + (1 - \lambda) Y\right) \leq \lambda \rho_t(X) + (1 - \lambda) \rho_t(Y);$$

normalization: $\rho_t(0) = 0$.

A process $(\rho_t)_{t \in [0,T]}$ is called dynamic convex risk measure for processes *if for each t, $\rho_t \colon \mathcal{R}^\infty_t \to L^\infty(\Omega, \mathcal{F}_t, \mathbb{P})$ is a conditional convex risk measure for processes.*

For $X \in \mathcal{R}^\infty$, we write

$$\rho_t(X) := \rho_t(\pi_{t,T}(X)).$$

Example 6.24 (dynamic entropic risk measure). *If we take the entropic risk measure for random variables,*

$$\widetilde{\rho}(X) = \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[e^{-\beta X} \right], \quad X \in \mathcal{X},$$

we can easily define the corresponding static entropic risk measure for processes via

$$\rho(X) = \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[e^{-\beta X_T} \right], \quad X \in \mathcal{R}^\infty.$$

The corresponding dynamic *entropic risk measure for processes is obtained by replacing the expectation by the conditional expectation:*

$$\rho_t(X) := \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[e^{-\beta X_T} \mid \mathcal{F}_t \right], \quad X \in \mathcal{R}^\infty.$$

This is a dynamic convex risk measure for processes.

Exercise 12. *Check that the dynamic entropic risk measure ρ_t is indeed conditional cash invariant.*

Definition 6.25. *A dynamic convex risk measure for processes is called* time consistent *if*

$$\rho_t(X) = \rho_t(X \mathbb{1}_{[t,s)} - \rho_s(X) \mathbb{1}_{[s,T]}), \quad \forall X \in \mathcal{R}^\infty, \ \forall t \in [0,T], s \in [t,T].$$

Lemma 6.26. *The dynamic entropic risk measure ρ_t is time consistent.*

Proof. Let $t \in [0,T]$. Then for $X \in \mathcal{R}^\infty$,

$$\begin{aligned}
&\rho_t(X \mathbb{1}_{[t,s)} - \rho_s(X) \mathbb{1}_{[s,T]}) \\
&= \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[e^{-\beta \left(X_T \mathbb{1}_{[t,s)}(T) - \rho_s(X) \mathbb{1}_{[s,T]}(T) \right)} | \mathcal{F}_t \right] \\
&= \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[e^{\beta \rho_s(X)} | \mathcal{F}_t \right] \\
&= \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[\exp \left(\beta \cdot \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[e^{-\beta X_T} | \mathcal{F}_s \right] \right) | \mathcal{F}_t \right] \\
&= \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[\mathbb{E}^{\mathbb{P}} \left[e^{-\beta X_T} | \mathcal{F}_s \right] | \mathcal{F}_t \right] \qquad \text{and as } s \geq t \\
&= \frac{1}{\beta} \ln \mathbb{E}^{\mathbb{P}} \left[e^{-\beta X_T} | \mathcal{F}_t \right] \\
&= \rho_t(X).
\end{aligned}$$

□

In order to better understand what time inconsistency means, let us look at an example of a time inconsistent risk measure: the dynamic Value at Risk.

A natural dynamic version of Value at Risk at level $\alpha \in (0,1)$ is

$$V@R_t^\alpha(X) = \operatorname{essinf} \left\{ m \in L^\infty(\mathcal{F}_t) \mid \ \mathbb{P}(m + X < 0 | \mathcal{F}_t) \leq \alpha \right\}.$$

In the following example we will see that $V@R_t^\alpha$ is time inconsistent and what this entails.

Example 6.27 (V@R is time inconsistent.[25]). *Fix an initial value $s_0 > 0$, a volatility $\sigma > 0$ and a constant $\nu \in \mathbb{R}$. Let Z_1, Z_2 be i.i.d. $\mathcal{N}(0,1)$-distributed random variables and define a stock price process $(S_t)_{t\in\{0,1,2\}}$ by*

$$S_0 = s_0 \quad \textit{and} \quad S_t = s_0 \exp\left(\sigma \sum_{k=1}^{t} Z_k + \nu t\right), \quad t \in \{1,2\}.$$

Let $(\mathcal{F}_t)_{t\in\{0,1,2\}}$ be the filtration generated by S and set

$$R_1 := \frac{S_1}{S_0}, \quad R_2 := \frac{S_2}{S_1}.$$

Choose an arbitrary probability level $\alpha \in (0,1)$. Then there exist constants a, b and $B > b$ such that

$$\begin{aligned}
&\alpha < \mathbb{P}(R_2 \leq b) < \mathbb{P}(R_2 \leq B),\\
&\mathbb{P}(R_1 \leq a)\mathbb{P}(R_2 \leq b) < \alpha < \mathbb{P}(R_1 \leq a)\mathbb{P}(R_2 \leq B).
\end{aligned}$$

We want to determine the dynamic Value at Risk for the random payoffs

$$\begin{aligned}
X &:= -C\mathbb{1}_E + d\mathbb{1}_{E^c} \quad \textit{and}\\
Y &:= -c\mathbb{1}_F + D\mathbb{1}_{F^c},
\end{aligned}$$

where C, c, D and d are constants such that $C > c > 0$ and $D > d > 0$ and E and F are the events given by

$$\begin{aligned}
E &:= \{R_1 \leq a, R_2 \leq b\},\\
F &:= \{R_1 \leq a, R_2 \leq B\}.
\end{aligned}$$

A possible choice for the numerous constants introduced above is the following:

$$s_0 = 1, \quad \sigma = 0.1, \quad \nu = 0.06 - \frac{\sigma^2}{2}, \quad \alpha = 0.05, \quad a = 1, \quad b = 0.95, \quad B = 1.$$

With this choice,

$$\begin{aligned}
E &= \{R_1 \leq 1, R_2 \leq 0.95\},\\
F &= \{R_1 \leq 1, R_2 \leq 1\}.
\end{aligned}$$

Let us calculate $V@R_t^\alpha$ for $t \in \{0,1\}$:

$$\begin{aligned}
V@R_0^\alpha(X) &= \operatorname{essinf}\{m \mid \mathbb{P}(m + X < 0) \leq \alpha\}\\
&= \operatorname{essinf}\{m \mid \mathbb{P}(-C\mathbb{1}_E + d\mathbb{1}_{E^c} < -m) \leq \alpha\}\\
&= -d
\end{aligned}$$

and analogously

$$V@R_0^\alpha(Y) = c.$$

[25] from: Example 3.1 in *Time-inconsistent VaR and time-consistent alternatives* (2008) by P. Cheridito & M. Stadje

For $t = 1$ we get

$$\begin{aligned} V@R_1^\alpha(X) &= \operatorname{essinf}\{m \in L^\infty(\mathcal{F}_1) \mid \mathbb{P}(m + X < 0|\mathcal{F}_1) \leq \alpha\} \\ &= \begin{cases} C, & \text{if } R_1 \leq a, \\ -d, & \text{if } R_1 > a. \end{cases} \end{aligned}$$

and analogously

$$V@R_1^\alpha(X) = \begin{cases} c, & \text{if } R_1 \leq a, \\ -D, & \text{if } R_1 > a. \end{cases}$$

Thus we can compare the risks of X and Y at each moment in time:

$$V@R_1^\alpha(X) > V@R_1^\alpha(Y) \quad \text{and} \quad V@R_0^\alpha(X) < V@R_0^\alpha(Y). \tag{6.11}$$

This means that $(V@R_0^\alpha, V@R_1^\alpha)$ is not time-consistent. Let us go a step further to see what this entails.
Assume that a trader wants to minimize $V@R_t^\alpha$ over the two available payoffs X and Y under the constraint that

$$\mathbb{E}[\text{payoff} \mid \mathcal{F}_t] \geq m_0.$$

If D and d are large enough such that $D > d \geq m_0$ and

$$-C\mathbb{P}(R_2 \leq b) + d\mathbb{P}(R_2 > b) \geq m_0 \quad \text{and} \quad -c\mathbb{P}(R_2 \leq B) + D\mathbb{P}(R_2 > B) \geq m_0,$$

then

$$\begin{aligned} \mathbb{E}[X \mid \mathcal{F}_1] &= \begin{cases} -C\mathbb{P}(R_2 \leq b) + d\mathbb{P}(R_2 > b), & \text{if } R_1 \leq a \\ d & \text{if } R_1 > a \end{cases} \\ &\geq m_0, \\ \mathbb{E}[Y \mid \mathcal{F}_1] &= \begin{cases} -c\mathbb{P}(R_2 \leq B) + D\mathbb{P}(R_2 > B), & \text{if } R_1 \leq a \\ D & \text{if } R_1 > a \end{cases} \\ &\geq m_0. \end{aligned}$$

Consequently, $\mathbb{E}[X] \geq m_0$ and $\mathbb{E}[Y] \geq m_0$, hence X and Y satisfy the constraint at all times. By (6.11), the risk of X is smaller than the risk of Y when perceived at time $t = 0$, hence the trader will prefer the future payoff X when asked at time $t = 0$. However, at time $t = 1$ he will regret his decision because at that time Y will have a lower risk than X.

A possible remedy to the problem of inconsistency in discrete time settings is by composing single-period risk measures in a time-consistent way. For Value at Risk, this could be done in the following recursive manner:

$$\begin{aligned} ComV@R_{T-1}^\alpha &:= V@R_{T-1}^\alpha, \\ ComV@R_t^\alpha &:= V@R_t^\alpha(-ComV@R_{t+1}^\alpha(X)), \quad t \in \{0, \dots, T-2\}. \end{aligned}$$

By construction $ComV@R^\alpha$ is time-consistent.

7 Backward Stochastic Differential Equations

This Section follows Chapter 6 in [Pha09].

7.1 General Properties

7.1.1 Existence and Uniqueness Results

Let $W = (W_t)t \in [0,T]$ be a standard d-dimensional BM on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, where the filtration is the natural filtration generated by W and $T < \infty$ is a fixed horizon. We introduce the following spaces[26]:

$$\begin{aligned}
\mathcal{P}_n(0,T) &:= \{\phi\colon \Omega \times [0,T] \to \mathbb{R}^n \mid \phi \text{ progressively measurable process}\} \\
\mathcal{S}^2(0,T) &:= \left\{ \phi \in \mathcal{P}_1(0,T) \;\middle|\; \mathbb{E}\left[\sup_{0\le t\le T} |\phi_t|^2\right] < \infty \right\} \\
\mathcal{H}^2_d(0,T) &:= \left\{ \phi \in \mathcal{P}_d(0,T) \;\middle|\; \mathbb{E}\left[\int_0^T |\phi_t|^2 \,\mathrm{d}t\right] < \infty \right\}
\end{aligned}$$

We want to consider the following backward stochastic differential equation (BSDE):

$$-\,\mathrm{d}Y_t = f(t, Y_t, Z_t)\mathrm{d}t - Z_t \cdot \mathrm{d}W_t, \quad Y_T = \xi, \tag{7.1}$$

where *driver* (or *generator*) f and *terminal condition* ξ shall satisfy the following conditions:

Assumption 7.1.

- $\xi \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}; \mathbb{R})$
- $f\colon \Omega \times [0,T] \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ *with short notation* $f(t,y,z)$ *(we suppress* ω*) is progressively measurable for all* $(y,z) \in \mathbb{R} \times \mathbb{R}^d$
- $f(t,0,0) \in \mathcal{H}^2_1(0,T)$
- f *satisfies a uniform Lipschitz condition, i.e., there exists a constant* C_f *such that* $\mathrm{d}\mathbb{P} \otimes \mathrm{d}t$ *a.e.*

$$|f(t,y_1,z_1) - f(t,y_2,z_2)| \le C_f(|y_1 - y_2| + |z_1 - z_2|), \quad \forall y_1, y_2 \in \mathbb{R}\, \forall z_1, z_2 \in \mathbb{R}^d$$

Definition 7.2. *A solution to the BSDE* (7.1) *with driver* f *and terminal condition* ξ *is a pair* $(Y,Z) \in \mathcal{S}^2(0,T) \times \mathcal{H}^2_d(0,T)$ *satisfying*

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)\mathrm{d}s - \int_t^T Z_s \cdot \mathrm{d}W_s, \qquad t \in [0,T].$$

Theorem 7.3. *Given a pair* (ξ, f) *satisfying Assumption 7.1, there exists a unique solution* (Y,Z) *to BSDE* (7.1).

[26] We follow the notation used in [Pha09] that is common in the BSDE literature. Observe that the space $\mathcal{H}^2_d(0,T)$ was denoted by $\mathcal{V}^d$ in Section 2.

For the proof of this theorem we need two results from Stochastic Analysis, which can be found e.g. as Theorem 1.1.4 and Theorem 1.1.6 in [Pha09].

Theorem 7.4 (Doob's maximal L^2-inequality). *Let $X = (X_t)_{t\in[0,t]}$ be a nonnegative right-continuous L^2-submartingale. Then for all stopping times τ with values in $[0, T]$ we have*

$$\mathbb{E}\left[\sup_{0\le t\le\tau} |X_t|^2\right] \le 4\mathbb{E}\left[X_\tau^2\right].$$

Let us quickly remark that there are more general versions of Doob's maximal inequality, but the above one is sufficient for our purpose. In particular, observe that the expectation on the LHS can be taken over $\sup_t |X_t|^2$ or $(\sup_t |X_t|)^2$ — in both cases the result is the same as long as we have non-negativity.

Theorem 7.5 (Burkholder-Davis-Gundy inequality). *For all $p > 0$ there exist constants c_p, C_p such that for all continuous local martingales $M = (M_t)_{t\in\mathbb{T}}$ and all stopping times τ with values in $\overline{\mathbb{T}}$, we have*

$$c_p\mathbb{E}\left[\langle M\rangle_\tau^{p/2}\right] \le \mathbb{E}\left[\sup_{0\le t<\tau} |M_t|\right]^p \le C_p\mathbb{E}\left[\langle M\rangle_\tau^{p/2}\right]$$

In particular, if (with $p = 1$) a continuous local martingale M satisfies $\mathbb{E}\left[\sqrt{\langle M\rangle_t}\right] < \infty$ for all $t \in \mathbb{T}$, then M is a martingale.

Proof of Theorem 7.3. The proof is based on a fixed point argument. Define a function Φ from $\mathcal{S}^2(0,T) \times \mathcal{H}_d^2(0,T)$ to itself via $\Phi(U,V) = (Y,Z)$ satisfying

$$Y_t = \xi + \int_t^T f(s, U_s, V_s)\mathrm{d}s - \int_t^T Z_s \cdot \mathrm{d}W_s. \tag{7.2}$$

The solution pair (Y,Z) is constructed as follows: Let

$$M_t := \mathbb{E}\left[\xi + \int_0^T f(s, U_s, V_s)\mathrm{d}s \,\middle|\, \mathcal{F}_t\right], \qquad t \in [0,T]. \tag{7.3}$$

Under Assumption 7.1 on ξ and f, M is a square integrable martingale. From the MRT (a multi-dimensional version of Theorem 2.36) we infer that there exists a unique $Z \in \mathcal{H}_d^2(0,T)$ such that

$$M_t = M_0 + \int_0^t Z_s \cdot \mathrm{d}W_s. \tag{7.4}$$

We define the process Y via

$$\begin{aligned}
Y_t &:= \mathbb{E}\left[\xi + \int_t^T f(s, U_s, V_s)\mathrm{d}s \,\middle|\, \mathcal{F}_t\right] \\
&\overset{(7.3)}{=} M_t - \int_0^t f(s, U_s, V_s)\mathrm{d}s \\
&\overset{(7.4)}{=} M_0 - \int_0^t f(s, U_s, V_s)\mathrm{d}s + \int_0^t Z_s \cdot \mathrm{d}W_s, \quad t \in [0,T],
\end{aligned}$$

thus $Y_T = \xi$ and

$$Y_t - \xi = Y_t - Y_T = \int_t^T f(s, U_s, V_s)\mathrm{d}s - \int_t^T Z_s \cdot \mathrm{d}W_s,$$

i.e., Y satisfies (7.2). Let us verify that

$$(Y, Z) \in \mathcal{S}^2(0, T) \times \mathcal{H}_d^2(0, T). \tag{7.5}$$

First, by Doob's maximal L^2-inequality,

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_t^T Z_s \cdot \mathrm{d}W_s\right|^2\right] \le 4\mathbb{E}\left[\int_0^T |Z_s|^2\,\mathrm{d}s\right] < \infty$$

which shows that the stochastic integral in (7.2) is in $\mathcal{S}^2(0, T)$. The other two summands in (7.2) are in $\mathcal{S}^2(0, T)$ as well by Assumption 7.1 (verifying this is left as an exercise), hence $Y \in \mathcal{S}^2(0, T)$. As $Z \in \mathcal{H}_d^2(0, T)$ by virtue of the MRT, we see that Φ is indeed a function from $\mathcal{S}^2(0, T) \times \mathcal{H}_d^2(0, T)$ into itself. Consequently, (Y, Z) is a solution to (7.1) if and only if it is a fixed point of Φ. To complete the proof, it is therefore sufficient to prove the following claim:

Claim: Φ admits a unique fixed point.

Proof of the claim: We want to show that Φ is a strict contraction on the Banach space $\mathcal{S}^2(0, T) \times \mathcal{H}_d^2(0, T)$ equipped with the norm

$$\|(Y, Z)\|_\beta := \left(\mathbb{E}\left[\int_0^T e^{\beta s}(|Y_s|^2 + |Z_s|^2)\mathrm{d}s\right]\right)^{1/2}$$

for a suitably chosen $\beta > 0$. To see this, let $(U, V), (U', V') \in \mathcal{S}^2(0, T) \times \mathcal{H}_d^2(0, T)$ and let $(Y, Z) = \Phi(U, V)$ and $(Y', Z') = \Phi(U', V')$. Set

$$\begin{aligned}
(\overline{U}, \overline{V}) &:= (U - U', V - V'),\\
(\overline{Y}, \overline{Z}) &:= (Y - Y', Z - Z'),\\
\overline{f}_t &:= f(t, U_t, V_t) - f(t, U_t', V_t').
\end{aligned}$$

If, for an arbitrary $\beta > 0$, we apply Itô's formula to $e^{\beta s}\overline{Y}_s^2$ for the interval $[0, T]$, then we obtain (with $\overline{Y}_T = 0$)

$$e^{\beta T}\overline{Y}_T^2 = 0 = \overline{Y}_0^2 + \int_0^T \beta e^{\beta s}\overline{Y}_s^2\mathrm{d}s + 2\int_0^T e^{\beta s}\overline{Y}_s\mathrm{d}\overline{Y}_s + \int_0^T e^{\beta s}\mathrm{d}\langle\overline{Y}\rangle_s,$$

hence

$$\begin{aligned}
\overline{Y}_0^2 &= -\int_0^T \beta e^{\beta s}\overline{Y}_s^2\mathrm{d}s - 2\int_0^T e^{\beta s}\overline{Y}_s\mathrm{d}\overline{Y}_s - \int_0^T e^{\beta s}\mathrm{d}\langle\overline{Y}\rangle_s\\
&= -\int_0^T \beta e^{\beta s}\overline{Y}_s^2\mathrm{d}s + 2\int_0^T e^{\beta s}\overline{Y}_s\overline{f}_s\mathrm{d}s - 2\int_0^T e^{\beta s}\overline{Y}_s\overline{Z}_s \cdot \mathrm{d}W_s - \int_0^T e^{\beta s}\left|\overline{Z}_s\right|^2\mathrm{d}s\\
&= -\int_0^T e^{\beta s}(\beta\overline{Y}_s^2 - 2\overline{Y}_s\overline{f}_s)\mathrm{d}s - \int_0^T e^{\beta s}\left|\overline{Z}_s\right|^2\mathrm{d}s - 2\int_0^T e^{\beta s}\overline{Y}_s\overline{Z}_s \cdot \mathrm{d}W_s.
\end{aligned} \tag{7.6}$$

Observe that

$$\mathbb{E}\left[\left(\int_0^T e^{2\beta t} Y_t^2 |Z_t|^2 \mathrm{d}t\right)^{1/2}\right] \leq \frac{e^{\beta T}}{2} \mathbb{E}\left[\sup_{0\leq t\leq T} Y_t^2 + \int_0^T |Z_t|^2 \mathrm{d}t\right] < \infty,$$

hence, from the Burkholder-Davis-Gundy inequality, we see that $\left(\int_0^t e^{\beta s} \overline{Y}_s \overline{Z}_s \cdot \mathrm{d}W_s\right)_t$ is a uniformly integrable martingale. By taking expectations in (7.6) (and by rearranging slightly), we get with the help of Young's inequality[27]

$$\begin{aligned}
&\mathbb{E}\left[\overline{Y}_0^2\right] + \mathbb{E}\left[\int_0^T e^{\beta s}\left(\beta \overline{Y}_s^2 + |\overline{Z}_s|^2\right) \mathrm{d}s\right] \\
&= 2\mathbb{E}\left[\int_0^T e^{\beta s} \overline{Y}_s \overline{f}_s \mathrm{d}s\right] \\
&\leq 2C_f \mathbb{E}\left[\int_0^T e^{\beta s} |\overline{Y}_s| (|\overline{U}_s| + |\overline{V}_s|) \mathrm{d}s\right] \qquad \text{by Ass. 7.1} \\
&= \mathbb{E}\left[\int_0^T 2\sqrt{2} C_f e^{\frac{\beta s}{2}} |\overline{Y}_s| \frac{1}{\sqrt{2}} e^{\frac{\beta s}{2}} (|\overline{U}_s| + |\overline{V}_s|) \mathrm{d}s\right] \\
&\overset{\text{Young}}{\leq} \mathbb{E}\left[\int_0^T \frac{1}{2}\left(2\sqrt{2} C_f e^{\frac{\beta s}{2}} |\overline{Y}_s|\right)^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_0^T \frac{1}{2}\left(\frac{1}{\sqrt{2}} e^{\frac{\beta s}{2}} (|\overline{U}_s| + |\overline{V}_s|)\right)^2 \mathrm{d}s\right] \\
&= 4C_f^2 \mathbb{E}\left[\int_0^T e^{\beta s} |\overline{Y}_s|^2 \mathrm{d}s\right] + \frac{1}{4}\mathbb{E}\left[\int_0^T e^{\beta s} (|\overline{U}_s| + |\overline{V}_s|)^2 \mathrm{d}s\right] \\
&\leq 4C_f^2 \mathbb{E}\left[\int_0^T e^{\beta s} |\overline{Y}_s|^2 \mathrm{d}s\right] + \frac{1}{2}\mathbb{E}\left[\int_0^T e^{\beta s}\left(|\overline{U}_s|^2 + |\overline{V}_s|^2\right) \mathrm{d}s\right],
\end{aligned}$$

where the last inequality uses that $(a+b)^2 \leq 2(a^2+b^2)$. Now choose $\beta := 4C_f^2 + 1$. This yields

$$\mathbb{E}\left[\int_0^T e^{\beta s}\left(|\overline{Y}_s|^2 + |\overline{Z}_s|^2\right) \mathrm{d}s\right] \leq \frac{1}{2}\mathbb{E}\left[\int_0^T e^{\beta s}\left(|\overline{U}_s|^2 + |\overline{V}_s|^2\right) \mathrm{d}s\right].$$

This inequality shows that Φ is a strict contraction on space suggested.

With the claim proven, the proof of Theorem 7.3 is also completed. □

7.1.2 Linear BSDEs

Let us consider the particular case where the generator f is linear in y and z, i.e., the BSDE is of the form

$$-\mathrm{d}Y_t = (A_t Y_t + Z_t \cdot B_t + C_t)\mathrm{d}t - Z_t \cdot \mathrm{d}W_t, \quad Y_T = \xi, \tag{7.7}$$

where A and B are bounded progressively measurable processes with values in $\mathbb{R}$ and $\mathbb{R}^d$, respectively, and $C \in \mathcal{H}_1^2(0,T)$. We can solve this BSDE explicitly.

[27] Young's inequality: $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ if $\frac{1}{p} + \frac{1}{q} = 1$ for $p, q > 0$.

Proposition 7.6. *The unique solution* (Y, Z) *to the linear BSDE* (7.7) *is given by*

$$\Gamma_t Y_t = \mathbb{E}\left[\Gamma_T \xi + \int_t^T \Gamma_s C_s \mathrm{d}s \,\middle|\, \mathcal{F}_t \right], \tag{7.8}$$

where Γ *is the solution to the linear SDE*

$$\mathrm{d}\Gamma_t = \Gamma_t (A_t \mathrm{d}t + B_t \cdot \mathrm{d}W_t), \quad \Gamma_0 = 1.$$

Proof. We apply the integration-by-parts formula (Theorem 2.33) to $\Gamma_t Y_t$ and obtain

$$\begin{aligned}
\mathrm{d}(\Gamma_t Y_t) &= \Gamma_t \mathrm{d}Y_t + Y_t \mathrm{d}\Gamma_t + \mathrm{d}\Gamma_t \mathrm{d}Y_t \\
&= -(A_t Y_t + Z_t \cdot B_t + C_t)\Gamma_t \mathrm{d}t + \Gamma_t Z_t \cdot \mathrm{d}W_t + \Gamma_t A_t Y_t \mathrm{d}t + \Gamma_t Y_t B_t \cdot \mathrm{d}W_t + \Gamma_t Z_t \cdot B_t \mathrm{d}t \\
&= -\Gamma_t C_t \mathrm{d}t + \Gamma_t (Z_t + Y_t B_t) \cdot \mathrm{d}W_t.
\end{aligned}$$

Consequently, with $\Gamma_0 = 1$,

$$\Gamma_t Y_t + \int_0^t \Gamma_s C_s \mathrm{d}s = Y_0 + \int_0^t \Gamma_s (Z_s + Y_s B_s) \cdot \mathrm{d}W_s. \tag{7.9}$$

As in the previous proof, we want to take expectations and for this we want to show that the stochastic integral is a martingale. To see this, observe that

$$\mathbb{E}\left[\sup_{t \in [0,T]} |\Gamma_t|^2 \right] < \infty,$$

since A and B are bounded by assumption. If we denote by b_∞ the upper bound of B, then we have (again by applying Young's inequality)

$$\begin{aligned}
\mathbb{E}\left[\left(\int_0^T \Gamma_s^2 |Z_s + Y_s B_s|^2 \mathrm{d}s \right)^{\frac{1}{2}} \right] &\le \mathbb{E}\left[\sup_t |\Gamma_t| \left(\int_0^T |Z_s + Y_s B_s|^2 \mathrm{d}s \right)^{\frac{1}{2}} \right] \\
&\le \frac{1}{2} \mathbb{E}\left[\sup_t |\Gamma_t|^2 + \int_0^T |Z_s + Y_s B_s|^2 \mathrm{d}s \right] \\
&\le \frac{1}{2} \mathbb{E}\left[\sup_t |\Gamma_t|^2 + \int_0^T 2(|Z_s|^2 + |Y_s B_s|^2) \mathrm{d}s \right] \\
&\le \frac{1}{2} \mathbb{E}\left[\sup_t |\Gamma_t|^2 + 2 \int_0^T |Z_s|^2 \mathrm{d}s + 2b_\infty^2 \int_0^T |Y_s|^2 \mathrm{d}s \right] \\
&< \infty.
\end{aligned}$$

Again with the Burkholder-Davis-Gundy inequality we can infer that the stochastic integral in (7.9) is not only a local martingale, but a uniformly integrable martingale. By adding Y_0, the RHS of (7.9) is a martingale. Thus, the LHS has the representation

$$\begin{aligned}
\Gamma_t Y_t + \int_0^t \Gamma_s C_s \mathrm{d}s &= \mathbb{E}\left[\Gamma_T Y_T + \int_0^T \Gamma_s C_s \mathrm{d}s \,\middle|\, \mathcal{F}_t \right] \\
&= \mathbb{E}\left[\Gamma_T \xi + \int_0^T \Gamma_s C_s \mathrm{d}s \,\middle|\, \mathcal{F}_t \right],
\end{aligned} \tag{7.10}$$

which proves (7.8). Once Y has been computed this way, Z is given via the martingale representation in (7.9) for the martingale in (7.10). □

7.1.3 Comparison principle

The following result will prove to be essential if one wants to optimize over a family of BSDEs. Regarding the notation, we emphasize that there are no squares in the theorem or proof, only upper indices.

Theorem 7.7. *Let (ξ^1, f^1) and (ξ^2, f^2) be pairs of terminal conditions and generators satisfying Assumption 7.1. Let (Y^1, Z^1) and (Y^2, Z^2) be the solutions to the corresponding BSDEs. Suppose that*

- $\xi^1 \leq \xi^2$ *a.s.;*
- $f^1(t, Y_t^1, Z_t^1) \leq f^2(t, Y_t^1, Z_t^1)\ \mathrm{d}\mathbb{P} \otimes \mathrm{d}t$*-a.e.;*
- $f^2(t, Y_t^1, Z_t^1) \in \mathcal{H}_1^2(0, T)$.

Then almost surely $Y_t^1 \leq Y_t^2$ for all $t \in [0, T]$. If, additionally, one has $Y_0^2 \leq Y_0^1$, then $Y_t^1 = Y_t^2$ for all $t \in [0, T]$. In particular, if $\mathbb{P}(\xi^1 < \xi^2) > 0$ or $f^1(t, \cdot, \cdot) < f^2(t, \cdot, \cdot)$ on a set of strictly positive measure $\mathrm{d}\mathbb{P} \otimes \mathrm{d}t$, then $Y_0^1 < Y_0^2$.

Proof. To simplify the notation, we assume $d = 1$. We define (careful with the differences to the notation from the proof of Theorem 7.3):

$$
\begin{aligned}
(\overline{Y}, \overline{Z}) &= (Y^2 - Y^1, Z^2 - Z^1),\\
\overline{f}_t &= f^2(t, Y_t^1, Z_t^1) - f^1(t, Y_t^1, Z_t^1),\\
\Delta_t^y &= \frac{f^2(t, Y_t^2, Z_t^2) - f^2(t, Y_t^1, Z_t^2)}{Y_t^2 - Y_t^1} \mathbb{1}_{\{Y_t^2 - Y_t^1 \neq 0\}},\\
\Delta_t^z &= \frac{f^2(t, Y_t^1, Z_t^2) - f^2(t, Y_t^1, Z_t^1)}{Z_t^2 - Z_t^1} \mathbb{1}_{\{Z_t^2 - Z_t^1 \neq 0\}}
\end{aligned}
$$

Then $(\overline{Y}, \overline{Z})$ satisfies the *linear* BSDE

$$
-\mathrm{d}\overline{Y}_t = (\Delta_t^y \overline{Y}_t + \Delta_t^z \overline{Z}_t + \overline{f}_t)\mathrm{d}t - \overline{Z}_t \mathrm{d}W_t, \qquad \overline{Y}_T = \xi^2 - \xi^1.
$$

Since f^2 is uniformly Lipschitz in y and z, the processes Δ^y and Δ^z are bounded. Moreover, the process $\overline{f}$ is in $\mathcal{H}_1^2(0, T)$. From Proposition 7.6 we know that $\overline{Y}$ is given by

$$
\Gamma_t \overline{Y}_t = \mathbb{E}\left[\Gamma_T(\xi^2 - \xi^1) + \int_t^T \Gamma_s \overline{f}_s \mathrm{d}s \,\middle|\, \mathcal{F}_t\right], \tag{7.11}
$$

where Γ is strictly positive. Hence, $\overline{Y}_t \geq 0$ for all $\in [0, T]$ almost surely. Furthermore, if $\overline{Y}_0 = 0$, then from (7.11) we infer that

$$
0 = \mathbb{E}\left[\Gamma_T(\xi^2 - \xi^1) + \int_0^T \Gamma_s \overline{f}_s \mathrm{d}s\right].
$$

With Γ being strictly positive and $\xi^2 - \xi^1$ and $\overline{f}$ at least non-negative by assumption, this implies that we have indeed $\xi^2 = \xi^1$ a.s. and $\overline{f} = 0$ a.e. Again with (7.11) we conclude that $\overline{Y}_t = 0$ for all $t \in [0, T]$ almost surely. $\square$

7.1.4 BSDEs, PDEs and Feynman-Kac formulae

Recall from Remark 5.9 (Feynman-Kac formula) that the solution to the linear PDE[28]

$$\frac{\partial v}{\partial t}(t,x) + \mathcal{L}v(t,x) + f(t,x) = 0, \qquad (t,x) \in [0,T) \times \mathbb{R}^n,$$
$$v(T,x) = g(x), \quad x \in \mathbb{R}^n,$$

with

$$\mathcal{L}v = \gamma(x) \cdot D_x v + \frac{1}{2}\operatorname{tr}(\sigma(x)\sigma^T(x)D_x^2 v)$$

has the representation

$$v(t,x) = \mathbb{E}\left[\int_t^T f(u, X_u^{t,x})\mathrm{d}u + g(X_T^{t,x})\right]$$

for a stochastic process satisfying SDE

$$\mathrm{d}X_s = \gamma(X_s)\mathrm{d}s + \sigma(X_s)\mathrm{d}W_s, \quad s \in [t,T], \quad X_t = x. \tag{7.12}$$

Let us now look at an extension of the Feynman-Kac formula to PDEs of the form

$$\frac{\partial v}{\partial t}(t,x) + \mathcal{L}v(t,x) + f(t,x,v,\sigma^T D_x v) = 0, \qquad (t,x) \in [0,T) \times \mathbb{R}^n, \tag{7.13}$$
$$v(T,x) = g(x), \quad x \in \mathbb{R}^n, \tag{7.14}$$

We want to represent the solution to this PDE by means of the BSDE

$$-\mathrm{d}Y_s = f(s, X_s, Y_s, Z_s)\mathrm{d}s - Z_s \cdot \mathrm{d}W_s, \quad s \in [t,T], \quad Y_T = g(X_T) \tag{7.15}$$

and the forward SDE (7.12).

The $\mathbb{R}^n$-valued γ and the $\mathbb{R}^{n\times d}$-valued σ satisfy a uniform Lipschitz condition; $f\colon [0,T] \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ satisfies a linear growth condition in (x,y,z) and a Lipschitz condition in (y,z), uniformly in (t,x); the continuous function g satisfies a linear growth condition. Under these conditions, the terminal condition and driver of BSDE (7.15) satisfy Assumption 7.1. (To verify this, one needs an estimate for the second moments of X.) By the Markov property of the diffusion X, we also have $Y_t = v(t, X_t)$ for

$$v(t,x) := Y_t^{t,x},$$

where $v\colon [0,T] \times \mathbb{R}^n \to \mathbb{R}$ is a deterministic function; $X_s^{t,x}$ $(s \in [t,T])$ is a solution to (7.12) starting from x at time t and $(Y_s^{t,x}, Z_s^{t,x})_{s\in[t,T]}$ is the solution to the BSDE (7.15) with $X_s = X_s^{t,x}$ $(s \in [t,T])$. This framework is called a Markovian case for the BSDE.

The following verification result for PDE (7.13) is analogous to the verification theorem for HJB equations, Theorem 5.7. If shows that a classical solution to the semilinear PDE provides a solution to the BSDE.

[28] In order to stick with the notation from [Pha09], we exchange the dimensions d and n compared to our choice in Section 5.2.

Proposition 7.8. *Let $v \in C^{1,2}([0,T]\times\mathbb{R}^n)\cap C^0([0,T]\times\mathbb{R}^n)$ be a classical solution to* (7.13)-(7.14)*, satisfying a linear growth condition and such that for some constants $C > 0$ and $q > 0$, $|D_x v(t,x)| \le C(1+|x|^q)$ for all $x \in \mathbb{R}^n$. Then the pair (Y, Z) defined by*

$$Y_t = v(t, X_t), \quad Z_t = \sigma^T(X_t) D_x v(t, X_t), \quad t \in [0,T],$$

is the solution to BSDE (7.15)

Proof. If we apply Itô's formula to $v(t, X_t)$, we get

$$\begin{aligned}
\mathrm{d}Y_t &= \mathrm{d}v(t, X_t)\\
&= \frac{\partial}{\partial t} v(t,X_t)\mathrm{d}t + D_x v(t,X_t)\cdot \mathrm{d}X_t + \frac{1}{2} D_x^2 v(t,X_t)\mathrm{d}\langle X\rangle_t\\
&= \frac{\partial}{\partial t} v(t,X_t)\mathrm{d}t + D_x v(t,X_t)\cdot (\gamma(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t) + \frac{1}{2}\operatorname{tr}(\sigma(X_t)\sigma^T(X_t)D_x^2 v(t,X_t))\mathrm{d}t\\
&= \left(\frac{\partial}{\partial t} v(t,X_t) + \gamma(X_t)\cdot D_x v(t,X_t) + \frac{1}{2}\operatorname{tr}(\sigma(X_t)\sigma^T(X_t)D_x^2 v(t,X_t))\right)\mathrm{d}t + \sigma^T(X_t)D_x v(t,X_t)\cdot \mathrm{d}W_t\\
&= \left(\frac{\partial}{\partial t} v(t,X_t) + \mathcal{L}v(t,X_t)\right)\mathrm{d}t + \sigma^T(X_t)D_x v(t,X_t)\cdot \mathrm{d}W_t\\
&\overset{(7.13)}{=} -f(t, X_t, v(t,X_t), \sigma^T(X_t)D_x v(t,X_t))\mathrm{d}t + \sigma^T(X_t)D_x v(t,X_t)\cdot \mathrm{d}W_t\\
&\overset{\text{def. } Y,Z}{=} -f(t, X_t, Y_t, Z_t)\mathrm{d}t + Z_t \cdot \mathrm{d}W_t.
\end{aligned}$$

Furthermore, $v(T, X_T) \overset{(7.14)}{=} g(X_T) = Y_T$. Finally, from the growth conditions on v and $D_x v$, we have that $(Y, Z) \in \mathcal{S}(0,T)\times\mathcal{H}_d^2(0,T)$, which completes the proof. □

Remark 7.9. *We (almost) have the converse statement: The function $v(t,x) = Y_t^{t,x}$ is continuous on $[0,T]\times\mathbb{R}^n$ and it is a* viscosity solution *to* (7.13)-(7.14)*. This is stated as Theorem 6.3.3 and proven in [Pha09].*

7.2 Control and BSDEs

The comparison principle, Theorem 7.7, states that if both the driver and the terminal condition of one BSDE lie below those of another BSDE, then the same holds true for the Y-component of the solutions of the BSDEs. Taking this even further, we now consider a family of BSDEs.

Theorem 7.10. *Let $\mathcal{A}$ be a subset of all progressively measurable processes. Let (ξ, f) and (ξ^α, f^α), $\alpha \in \mathcal{A}$, be a family of pairs of terminal conditions and drivers, each pair satisfying Assumption 7.1. Let (Y, Z) and (Y^α, Z^α), $\alpha \in \mathcal{A}$, be the solutions to the corresponding BSDEs. Suppose that there exists $\widehat{\alpha} \in \mathcal{A}$ such that*

$$\begin{aligned}
f(t, Y_t, Z_t) = \operatorname{essinf}_{\alpha\in\mathcal{A}} f^\alpha(t, Y_t, Z_t) = f^{\widehat{\alpha}}(t, Y_t, Z_t), \quad \mathrm{d}\mathbb{P}\otimes\mathrm{d}t - \textit{a.e.},\\
\xi = \operatorname{essinf}_{\alpha\in\mathcal{A}} \xi^\alpha = \xi^{\widehat{\alpha}}.
\end{aligned}$$

Then

$$Y_t = \operatorname{essinf}_{\alpha\in\mathcal{A}} Y_t^\alpha = Y_t^{\widehat{\alpha}}, \quad \forall\, t \in [0,T],\ \mathbb{P} - \textit{a.s.}$$

Proof. From the comparison principle (Theorem 7.7), since $\xi \leq \xi^\alpha$ and $f(t, Y_t, Z_t) \leq f^\alpha(t, Y_t, Z_t)$, we have $Y_t \leq Y_t^\alpha$ for all $t \in [0, T]$, $\mathbb{P}$-a.s., for all $\alpha \in \mathcal{A}$. Hence,

$$Y_t \leq \operatorname{essinf}_{\alpha \in \mathcal{A}} Y_t^\alpha.$$

Moreover, if $\xi = \xi^{\widehat{\alpha}}$ and $f(t, Y_t, Z_t) = f^{\widehat{\alpha}}(t, Y_t, Z_t)$, then (Y, Z) and $(Y^{\widehat{\alpha}}, Z^{\widehat{\alpha}})$ are both solutions to the BSDE with driver $f^{\widehat{\alpha}}$ and terminal condition $\xi^{\widehat{\alpha}}$, hence, by uniqueness, these solutions coincide. This implies for the Y-component that

$$\operatorname{essinf}_{\alpha \in \mathcal{A}} Y_t^\alpha \leq Y_t^{\widehat{\alpha}} = Y_t \leq \operatorname{essinf}_{\alpha \in \mathcal{A}} Y_t^\alpha,$$

which completes the proof. □

In Section 5 we have seen how to solve a stochastic optimal control problem with the help of the dynamic programming principle. (The latter was used to derive the HJB equation.) Here we want to look at an alternative approach, called *Pontryagin's maximum principle*.

The framework is the following: Let X be a controlled diffusion process on $\mathbb{R}^n$ governed by

$$\mathrm{d}X_t = \gamma(X_t, \alpha_t)\mathrm{d}t + \sigma(X_t, \alpha_t)\mathrm{d}W_t, \quad X_0 = x, \tag{7.16}$$

where W is a d-dimensional standard BM and $\alpha \in \mathcal{A}$, the control process, is a progressively measurable process with values in A. The gain functional that is to be maximized is given by

$$J(\alpha) = \mathbb{E}\left[\int_0^T f(t, X_t, \alpha_t)\mathrm{d}t + g(X_T)\right],$$

where $f\colon [0, T] \times \mathbb{R}^n \times A \to \mathbb{R}$ is continuous in (t, x) for all $a \in A$; $g \in C^1(\mathbb{R}^n; \mathbb{R})$ is concave; and f and g satisfy a quadratic growth condition in x.

Remark 7.11. *Observe that, while J in Section 5 was a function of t, x and the control, here we only have a dependency on the control. The reason is that we do not apply the DPP, hence there is no need for such a dependency.*

We define the *generalized Hamiltonian* $\mathcal{H}\colon [0, T] \times \mathbb{R}^n \times A \times \mathbb{R}^n \times \mathbb{R}^{n \times d} \to \mathbb{R}$ by

$$\mathcal{H}(t, x, a, y, z) := \gamma(x, a) \cdot y + \operatorname{tr}(\sigma^T(x, a)z) + f(t, x, a). \tag{7.17}$$

There is a slight difference to the previous subsection: Now $y \in \mathbb{R}^n$ and $z \in \mathbb{R}^{n \times d}$, whereas before we had $Y \in \mathbb{R}$ and $Z \in \mathbb{R}^d$.

Remark 7.12. *So as to avoid confusion, observe the difference between the above defined* generalized Hamiltonian *and the* Hamiltonian *as it was defined in Section 5:*

$$H(t, x, p, M) := \sup_{a \in A}\left[\gamma(x, a) \cdot p + \frac{1}{2}\operatorname{tr}(\sigma(x, a)\sigma^T(x, a)M + f(t, x, a)\right].$$

We will assume that $\mathcal{H}$ is differentiable in x with derivative $D_x\mathcal{H}$. For each $\alpha \in \mathcal{A}$, we consider the following BSDE, which is also called the *adjoint equation*:

$$-\mathrm{d}Y_t = D_x\mathcal{H}(t, X_t, \alpha_t, Y_t, Z_t)\mathrm{d}t - Z_t\mathrm{d}W_t, \quad Y_T = D_xg(X_T). \tag{7.18}$$

Theorem 7.13. *Let $\widehat{\alpha} \in \mathcal{A}$ and let $\widehat{X}$ be the associated controlled diffusion. Suppose that there exists a solution $(\widehat{Y}, \widehat{Z})$ to the associated BSDE* (7.18) *such that*

$$\mathcal{H}(t, \widehat{X}_t, \widehat{\alpha}_t, \widehat{Y}_t, \widehat{Z}_t) = \max_{a \in A} \mathcal{H}(t, \widehat{X}_t, a, \widehat{Y}_t, \widehat{Z}_t), \quad t \in [0,T], \text{a.s.} \tag{7.19}$$

Suppose furthermore that for all $t \in [0,T]$,

$$(x, a) \mapsto \mathcal{H}(t, x, a, \widehat{Y}_t, \widehat{Z}_t) \text{ is a concave function.} \tag{7.20}$$

Then $\widehat{\alpha}$ is an optimal control, i.e.

$$J(\widehat{\alpha}) = \sup_{\alpha \in \mathcal{A}} J(\alpha).$$

Proof. For any $\alpha \in \mathcal{A}$, we have

$$J(\widehat{\alpha}) - J(\alpha) = \mathbb{E}\left[\int_0^T \Big(f(t, \widehat{X}_t, \widehat{\alpha}_t) - f(t, X_t, \alpha_t)\Big)\, \mathrm{d}t + \Big(g(\widehat{X}_T) - g(X_T)\Big)\right]. \tag{7.21}$$

Let us look at both summands inside of the expectation separately. First, by concavity of g and by Itô's integration-by-parts formula, we have, as $X_0 = \widehat{X}_0$,

$$\begin{aligned}
&\mathbb{E}\left[g(\widehat{X}_T) - g(X_T)\right] \\
&\geq \mathbb{E}\left[(\widehat{X}_T - X_T) \cdot D_x g(\widehat{X}_T)\right] \\
&\overset{(7.18)}{=} \mathbb{E}\left[(\widehat{X}_T - X_T) \cdot \widehat{Y}_T\right] \\
&\overset{\text{Itô}}{=} \mathbb{E}\left[\int_0^T (\widehat{X}_t - X_t) \cdot \mathrm{d}\widehat{Y}_t + \int_0^T \widehat{Y}_t \cdot (\mathrm{d}\widehat{X}_t - \mathrm{d}X_t) + \int_0^T \operatorname{tr}\left[(\sigma(\widehat{X}_t, \widehat{\alpha}_t) - \sigma(X_t, \alpha_t))^T \widehat{Z}_t\right] \mathrm{d}t\right] \\
&= \mathbb{E}\left[\int_0^T (\widehat{X}_t - X_t) \cdot (-D_x \mathcal{H}(t, \widehat{X}_t, \widehat{\alpha}_t, \widehat{Y}_t, \widehat{Z}_t))\mathrm{d}t + \int_0^T Y_t \cdot (\gamma(\widehat{X}_t, \widehat{\alpha}_t) - \gamma(X_t, \alpha_t))\mathrm{d}t\right] \\
&\quad + \mathbb{E}\left[\int_0^T \operatorname{tr}\left[(\sigma(\widehat{X}_t, \widehat{\alpha}_t) - \sigma(X_t, \alpha_t))^T \widehat{Z}_t\right] \mathrm{d}t\right].
\end{aligned} \tag{7.22}$$

Second, by definition of $\mathcal{H}$ we have

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T \Big(f(t, \widehat{X}_t, \widehat{\alpha}_t) - f(t, X_t, \alpha_t)\Big)\, \mathrm{d}t\right] \\
&= \mathbb{E}\left[\int_0^T \Big(\mathcal{H}(t, \widehat{X}_t, \widehat{\alpha}_t, \widehat{Y}_t, \widehat{Z}_t) - \mathcal{H}(t, X_t, \alpha_t, \widehat{Y}_t, \widehat{Z}_t)\Big)\, \mathrm{d}t\right] \\
&\quad - \mathbb{E}\left[\int_0^T \Big(\gamma(\widehat{X}_t, \widehat{\alpha}_t) - \gamma(X_t, \alpha_t)\Big) \cdot \widehat{Y}_t \mathrm{d}t + \int_0^T \operatorname{tr}\left[(\sigma(\widehat{X}_t, \widehat{\alpha}_t) - \sigma(X_t, \alpha_t))^T \widehat{Z}_t\right] \mathrm{d}t\right].
\end{aligned} \tag{7.23}$$

By adding (7.22) and (7.23) and replacing the sum in (7.21) we obtain

$$\begin{aligned}
J(\widehat{\alpha}) - J(\alpha) \geq\ & \mathbb{E}\left[\int_0^T \Big(\mathcal{H}(t, \widehat{X}_t, \widehat{\alpha}_t, \widehat{Y}_t, \widehat{Z}_t) - \mathcal{H}(t, X_t, \alpha_t, \widehat{Y}_t, \widehat{Z}_t)\Big)\, \mathrm{d}t\right] \\
&+ \mathbb{E}\left[\int_0^T (\widehat{X}_t - X_t) \cdot (-D_x \mathcal{H}(t, \widehat{X}_t, \widehat{\alpha}_t, \widehat{Y}_t, \widehat{Z}_t))\mathrm{d}t\right].
\end{aligned}$$

Under conditions (7.19) and (7.20), the integrands (taken together) are nonnegative, which completes the proof. □

As we have already compared this setting with that in Section 5, we now want to provide the connection between the maximum principle and the DPP.
The value function of the stochastic control problem considered above is

$$v(t,x) = \sup_{\alpha \in \mathcal{A}} \mathbb{E}\left[\int_t^T f(s, X_s^{t,x}, \alpha_s)\mathrm{d}s + g(X_T^{t,x})\right], \tag{7.24}$$

where $(X_s^{t,x})_{s\in[t,T]}$ is the solution to (7.16) starting from x at time t. The associated HJB equation is

$$\frac{\partial v}{\partial t} + \sup_{a\in A}\left[\mathcal{G}(t,x,a,D_x v, D_x^2 v)\right] = 0, \tag{7.25}$$

where for $(t,x,a,p,M) \in [0,T] \times \mathbb{R}^n \times A \times \mathbb{R}^n \times \mathrm{Sym}_n$ (if we denote by Sym_n the symmetric $n\times n$-matrices),

$$\mathcal{G}(t,x,a,p,M) := \gamma(x,a)\cdot p + \frac{1}{2}\operatorname{tr}\left[\sigma(x,a)\sigma(x,a)^T M\right] + f(t,x,a). \tag{7.26}$$

Theorem 7.14. *Suppose that $v \in C^{1,3}([0,T)\times\mathbb{R}^n) \cap C^0([0,T]\times\mathbb{R}^n)$ and that there exists an optimal control $\widehat{\alpha} \in \mathcal{A}$ to* (7.24) *with associated diffusion $\widehat{X}$. Then*

$$\mathcal{G}(t,\widehat{X}_t,\widehat{\alpha}_t, D_x v(t,\widehat{X}_t), D_x^2 v(t,\widehat{X}_t)) = \max_{a\in A}\mathcal{G}(t,\widehat{X}_t, a, D_x v(t,\widehat{X}_t), D_x^2 v(t,\widehat{X}_t)) \tag{7.27}$$

and the pair

$$(\widehat{Y}_t, \widehat{Z}_t) := (D_x v(t,\widehat{X}_t), D_x^2 v(t,\widehat{X}_t)\sigma(\widehat{X}_t,\widehat{\alpha}_t)) \tag{7.28}$$

is the solution to the adjoint BSDE (7.18).

Proof. Since $\widehat{\alpha}$ is an optimal control, we have[29]

$$\begin{aligned} v(t,\widehat{X}_t) &= \mathbb{E}\left[\int_t^T f(s,\widehat{X}_s,\widehat{\alpha}_s)\mathrm{d}s + g(\widehat{X}_T)\,\middle|\,\mathcal{F}_t\right] \\ &= -\int_0^t f(s,\widehat{X}_s,\widehat{\alpha}_s)\mathrm{d}s + M_t, \quad t\in[0,T], \text{ a.s.}, \end{aligned} \tag{7.29}$$

where M is the martingale

$$M_t = \mathbb{E}\left[\int_0^T f(s,\widehat{X}_s,\widehat{\alpha}_s)\mathrm{d}s + g(\widehat{X}_T)\,\middle|\,\mathcal{F}_t\right].$$

By applying Itô's formula to $v(t,\widehat{X}_t)$ we get

$$\begin{aligned} &\mathrm{d}v(t,\widehat{X}_t) \\ &= \frac{\partial}{\partial t}v(t,\widehat{X}_t)\mathrm{d}t + D_x v(t,\widehat{X}_t)\cdot \mathrm{d}\widehat{X}_t + \frac{1}{2}D_x^2 v(t,\widehat{X}_t)\mathrm{d}\langle\widehat{X}\rangle_t \\ &= \frac{\partial}{\partial t}v(t,\widehat{X}_t)\mathrm{d}t + D_x v(t,\widehat{X}_t)\cdot(\gamma(\widehat{X}_t,\widehat{\alpha}_t)\mathrm{d}t + \sigma(\widehat{X}_t,\widehat{\alpha}_t)\mathrm{d}W_t) + \frac{1}{2}\operatorname{tr}((\sigma\sigma^T)(\widehat{X}_t,\widehat{\alpha}_t)D_x^2 v(t,\widehat{X}_t))\mathrm{d}t \\ &= \left[\frac{\partial}{\partial t}v(t,\widehat{X}_t) + \mathcal{G}(t,\widehat{X}_t,\widehat{\alpha}_t,D_x v(t,\widehat{X}_t), D_x^2 v(t,\widehat{X}_t))\right]\mathrm{d}t + D_x v(t,\widehat{X}_t)\cdot\sigma(\widehat{X}_t)\mathrm{d}W_t \\ &\quad - f(t,\widehat{X}_t,\widehat{\alpha}_t)\mathrm{d}t \end{aligned}$$

[29] Observe that the second argument of v is $\widehat{X}_t$ and not its starting value, hence the conditional expectation.

Furthermore, from (7.29), we have that $\mathrm{d}v(t,\widehat{X}_t) = -f(t,\widehat{X}_t,\widehat{\alpha}_t)\mathrm{d}t + \mathrm{d}M_t$. From the MRT, M has a representation as stochastic integral, hence the $\mathrm{d}t$-terms in the two representations of $\mathrm{d}v(t,\widehat{X}_t)$ can be identified, giving us

$$\frac{\partial}{\partial t}v(t,\widehat{X}_t) + \mathcal{G}(t,\widehat{X}_t,\widehat{\alpha}_t,D_x v(t,\widehat{X}_t),D_x^2 v(t,\widehat{X}_t)) = 0. \tag{7.30}$$

Since v is smooth, it satisfies (7.25), which yields (7.27).

From (7.25) and (7.30), we have

$$\begin{aligned} 0 &= \frac{\partial}{\partial t}v(t,\widehat{X}_t) + \mathcal{G}(t,\widehat{X}_t,\widehat{\alpha}_t,D_x v(t,\widehat{X}_t),D_x^2 v(t,\widehat{X}_t)) \\ &\geq \frac{\partial}{\partial t}v(t,x) + \mathcal{G}(t,x,\widehat{\alpha}_t,D_x v(t,x),D_x^2 v(t,x)), \qquad \forall x \in \mathbb{R}^n. \end{aligned}$$

Since $v \in C^{1,3}$, the optimality condition for the above relation implies

$$\frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}v(t,x) + \mathcal{G}(t,x,\widehat{\alpha}_t,D_x v(t,x),D_x^2 v(t,x))\right)\bigg|_{x=\widehat{X}_t} = 0.$$

Recall the expressions for $\mathcal{H}$ (in Equation (7.17)) and $\mathcal{G}$ (in Equation (7.26)). With these, the previous equality can be rewritten as

$$\begin{aligned} 0 = {} & \frac{\partial^2 v}{\partial t \partial x}(t,\widehat{X}_t) + D_x^2 v(t,\widehat{X}_t)\gamma(\widehat{X}_t,\widehat{\alpha}_t) + \frac{1}{2}\,\mathrm{tr}(\sigma(\widehat{X}_t,\widehat{\alpha}_t)\sigma^T(\widehat{X}_t,\widehat{\alpha}_t)D_x^3 v(t,\widehat{X}_t)) \\ & + D_x\mathcal{H}(t,\widehat{X}_t,\widehat{\alpha}_t,D_x v(t,\widehat{X}_t),D_x^2 v(t,\widehat{X}_t)\sigma(\widehat{X}_t,\widehat{\alpha}_t)) \end{aligned} \tag{7.31}$$

By applying Itô's formula to $D_x v(t,\widehat{X}_t)$ we get

$$\begin{aligned} -\mathrm{d}\widehat{Y}_t &= -\mathrm{d}D_x v(t,\widehat{X}_t) \\ &= -\left[\frac{\partial^2 v}{\partial t \partial x}(t,\widehat{X}_t) + D_x^2 v(t,\widehat{X}_t)\gamma(\widehat{X}_t,\widehat{\alpha}_t) + \frac{1}{2}\,\mathrm{tr}(\sigma(\widehat{X}_t,\widehat{\alpha}_t)\sigma^T(\widehat{X}_t,\widehat{\alpha}_t)D_x^3 v(t,\widehat{X}_t))\right]\mathrm{d}t \\ &\quad - D_x^2 v(t,\widehat{X}_t)\sigma(\widehat{X}_t,\widehat{\alpha}_t)\mathrm{d}W_t \\ &\overset{(7.31)}{=} D_x\mathcal{H}(t,\widehat{X}_t,\widehat{\alpha}_t,D_x v(t,\widehat{X}_t),D_x^2 v(t,\widehat{X}_t)\sigma(\widehat{X}_t,\widehat{\alpha}_t)) - D_x^2 v(t,\widehat{X}_t)\sigma(\widehat{X}_t,\widehat{\alpha}_t)\mathrm{d}W_t \\ &= D_x\mathcal{H}(t,\widehat{X}_t,\widehat{\alpha}_t,\widehat{Y}_t,\widehat{Z}_t) - \widehat{Z}_t\mathrm{d}W_t \end{aligned}$$

Finally, since $v(T,\cdot) = g(\cdot)$, we have

$$D_x v(T;\widehat{X}_T) = D_x g(\widehat{X}_T),$$

which proves that $(\widehat{Y},\widehat{Z})$ is indeed the solution to BSDE (7.18). $\square$

7.3 BSDEs and risk measures

This subsection is based on parts of Chapter 3 in [Car09].

In Example 6.24 we have introduced the *dynamic entropic risk measure* as

$$\rho_t(X) := \frac{1}{\beta} \ln \mathbb{E}\left[e^{-\beta X_T} \mid \mathcal{F}_t\right], \quad X \in \mathcal{R}^\infty,$$

where $\mathcal{R}^\infty$ denotes the adapted càdlàg processes that are essentially bounded.
In the following result we want to link the dynamic entropic risk measure for a bounded ($\mathcal{F}_T$-measurable) random variable X_T to a BSDE. We assume that W is a d-dimensional BM and $(\mathcal{F}_t)_t$ its natural (augmented) filtration.

Proposition 7.15 (Proposition 3.12 in [Car09]). *Let $X_T \in L^\infty(\mathcal{F}_T)$. The dynamic entropic risk measure $(\rho_t)_{t\in[0,T]}$ for parameter $\beta > 0$ is a solution to the BSDE with driver $f(z) = \frac{\beta}{2}\|z\|^2$ and terminal condition $-X_T$:*

$$-\,\mathrm{d}\rho_t(X_T) = \frac{\beta}{2}\|Z_t\|^2 \,\mathrm{d}t - Z_t \cdot \mathrm{d}W_t, \quad \rho_T(X_T) = -X_T. \tag{7.32}$$

Remark 7.16. *The driver f is not uniformly Lipschitz continuous, but one can show that BSDEs with a quadratic driver have a solution nonetheless, provided the terminal condition is in L^∞ (instead of our usual requirement that it belongs to L^2). We also require the solution to be bounded. However, the solution to a quadratic BSDE is not necessarily unique. If such a solution (Y^*, Z^*) satisfies $Y^* \geq Y$ (resp. $Y^* \leq Y$) for all solutions (Y, Z), then (Y^*, Z^*) is called a* maximal *(resp.* minimal*)* solution*.*

Proof. Denote by M the process

$$M_t(X_T) = \mathbb{E}\left[\exp(-\beta X_T) \mid \mathcal{F}_t\right], \quad t \in [0, T].$$

By construction, M is a positive L^2-martingale. From the MRT, there exists a unique $g \in \mathcal{V}$ such that

$$\mathrm{d}M_t = g_t \cdot \mathrm{d}W_t. \tag{7.33}$$

Observe that if W is multidimensional, then so are Z and g and a multidimensional version[30] of the MRT applies. From the positivity of M and of β, we can rewrite (7.33) as

$$\mathrm{d}M_t = \beta M_t (Z_t \cdot \mathrm{d}W_t)$$

by letting $Z_t = \frac{g_t}{\beta M_t}$. By applying Itô's formula to $\frac{1}{\beta}\ln(M_t) = \rho_t(X)$ we get

$$\begin{aligned}
-\mathrm{d}\left(\rho_t(X)\right) &= -\mathrm{d}\left(\frac{1}{\beta}\ln(M_t)\right) \\
&= -\frac{1}{\beta M_t}\mathrm{d}M_t + \frac{1}{2\beta M_t^2}\mathrm{d}\langle M\rangle_t \\
&= -Z_t \cdot \mathrm{d}W_t + \frac{\beta}{2}\|Z_t\|^2 \,\mathrm{d}t.
\end{aligned}$$

[30] For the multi-dimensional version of the MRT, see e.g. Theorem 5.4.2 in *Stochastic Calculus II* by Steven E. Shreve.

Taking into consideration that, by construction, $\rho_T(X_T) = -X_T$, this shows that ρ_t satsifies BSDE (7.32). Finally, if X_T is bounded, then so is $\rho_t(X_T)$ and therefore also

$$\frac{\beta}{2}\mathbb{E}\left[\int_t^T |Z_s|^2\,\mathrm{d}s \,\middle|\, \mathcal{F}_t\right] = \mathbb{E}\left[\rho_t(X_T) - X_T \,|\, \mathcal{F}_t\right].$$

Conversely, if (7.32) has a solution (Y, Z) such that Y_T and $\mathbb{E}\left[\int_t^T |Z_s|^2\,\mathrm{d}s \,\middle|\, \mathcal{F}_t\right]$ are bounded, then so is Y. □

Remark 7.17. *As the drivers in [Car09] are denoted by g, the dynamic operator $\mathcal{Y}^g$, which to each terminal condition ξ_T associates the maximal solution to the BSDE with driver g and terminal condition ξ_T, is called g-*dynamic operator *and $\mathcal{Y}_0^g$ is called g-*expectation. *If we let (Y, Z) denote the solution to a BSDE with driver g and terminal condition $Y_T = \xi_T$, then $\mathbb{E}^g\left[\xi_T \,|\, \mathcal{F}_t\right] := Y_t$ is the g-conditional expectation of ξ_T. Contrary to the usual conditional expectation, this one is* non-linear. *We will continue denoting the driver by f.*

Consider a BSDE

$$-\,\mathrm{d}Y_t = f(t, Y_t, Z_t)\mathrm{d}t - Z_t \cdot \mathrm{d}W_t, \quad Y_T = -\xi, \tag{7.34}$$

whose drivers shall be such that there exists a unique (possibly maximal) solution. Denote by $\mathcal{Y}^f$ the operator such that $\mathcal{Y}_t^f(\xi) = Y_t$ is the (maximal) solution to the BSDE with driver f and terminal condition $-\xi$. Let us see which conditions on the driver guarantee that $\mathcal{Y}^f$ has the properties of a dynamic risk measure.

Proposition 7.18. *Let $\mathcal{Y}^f$ be the dynamic operator introduced above.*

1. *$\mathcal{Y}^f$ is monotonic, i.e., if $\xi \geq \xi'$ a.s., then $\mathcal{Y}_t^f(\xi) \leq \mathcal{Y}_t^f(\xi')$ for all $t \in [0, T]$, a.s.*

2. *$\mathcal{Y}^f$ is time-consistent, i.e. for $s \leq t \leq T$, $\mathcal{Y}_s^f(\xi) = \mathcal{Y}_s^f(-\mathcal{Y}_t^f(\xi))$ a.s.*

3. *If f is independent of y, then $\mathcal{Y}^f$ is translation invariant, i.e., $\mathcal{Y}_t^f(\xi + \eta_t) = \mathcal{Y}_t^f(\xi) - \eta_t$ a.s. for $t \in [0, T]$ and $\eta_t \in \mathcal{F}_t$.*

4. *If f is convex, then so is $\mathcal{Y}^f$, i.e., for any $t \in [0, T]$ $\lambda \in [0, 1]$ and for any terminal conditions ξ, ξ',*

$$\mathcal{Y}_t^f(\lambda\xi + (1-\lambda)\xi') \leq \lambda\mathcal{Y}_t^f(\xi) + (1-\lambda)\mathcal{Y}_t^f(\xi').$$

5. *If $f^1 \leq f^2$, then $\mathcal{Y}^{f^1} \leq \mathcal{Y}^{f^2}$.*

Let us remark that Statement 3 is actually an “iff”-statement, but for our purpose, one direction is sufficient.

Proof. Statements 1 and 5 follow immediately from the comparison principle. For the time consistency (Statement 2), also called *flow property*, observe that for $s \leq t \leq T$, if we denote by $\mathcal{Y}_u^f(r,\xi)$ the dynamic operator with driver f, applied to terminal condition ξ at terminal time r, then

$$\begin{aligned}
\mathcal{Y}_s^f(t,-\mathcal{Y}_t^f(T,\xi)) &= \mathcal{Y}_s\left(t,\xi-\int_t^T f(...)\mathrm{d}u+\int_t^T Z_u\mathrm{d}W_u\right)\\
&= -\xi+\int_t^T f(...)\mathrm{d}u-\int_t^T Z_u\mathrm{d}W_u+\int_s^t f(...)\mathrm{d}u-\int_s^t Z_u\mathrm{d}W_u\\
&= -\xi+\int_s^T f(...)\mathrm{d}u-\int_s^T Z_u\mathrm{d}W_u\\
&= \mathcal{Y}_s(T,\xi).
\end{aligned}$$

For the translation invariance (Statement 3), observe that for any $t\in[0,T]$ and $\eta_t\in\mathcal{F}_t$,

$$\begin{aligned}
\mathcal{Y}_t^f(\xi+\eta_t) &= -\xi+\int_t^T f(...)\mathrm{d}u-\int_t^T Z_u\mathrm{d}W_u-\eta_t\\
&= \mathcal{Y}_t^f(\xi)-\eta_t.
\end{aligned}$$

Finally, for the proof of convexity (Statement 4), let us consider two BSDE with parameters (ξ^1,f) and (ξ^2,f), respectively and their corresponding solutions (Y^i,Z^i) for $i=1,2$. Let $\lambda\in[0,1]$ and define $\widetilde{Y}_t:=\lambda Y_t^1+(1-\lambda)Y_t^2$ and $\widetilde{Z}_t$ accordingly. Then $\widetilde{Y}$ satisfies

$$\begin{aligned}
-\mathrm{d}\widetilde{Y}_t &= (\lambda f(t,Y_t^1,Z_t^1)+(1-\lambda)f(t,Y_t^2,Z_t^2))\mathrm{d}t-(\lambda Z_t^1+(1-\lambda)Z_t^2)\mathrm{d}W_t,\\
\widetilde{Y}_T &= \lambda\xi^1+(1-\lambda)\xi^2.
\end{aligned}$$

Since f is convex, we can rewrite the first equation as

$$-\mathrm{d}\widetilde{Y}_t=(f(t,\widetilde{Y}_t,\widetilde{Z}_t)+\alpha(t,Y_t^1,Y_t^2,Z_t^1,Z_t^2,\lambda))\mathrm{d}t-\widetilde{Z}_t\mathrm{d}W_t,$$

where α is an almost surely nonnegative process. Thus, by 5., the solution Y to the BSDE with parameters $(\lambda\xi^1+(1-\lambda)\xi^2,f)$ is for any $t\in[0,T]$ a.s. smaller than the solution $\widetilde{Y}$ to the above BSDE. □

7.3.1 A model with equilibrium pricing – not relevant for the exam

Let us look at the model presented in [HPDR10] and extended in [BLDR17].

Assume that there are two sources of randomness, represented by a 2-dimensional standard BM $W=(W^S,W^R)$ on a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in[0,T]},\mathbb{P})$, where the filtration is generated by W and augmented by the $\mathbb{P}$-nullsets. We do not observe W directly, but we observe the temperature that evolves according to

$$\mathrm{d}R_t=\mu_t^R\mathrm{d}t+b\mathrm{d}W_t^R,\quad R_0=r_0$$

and a stock price process that evolves according to

$$\mathrm{d}S_t=\mu_t^S S_t\mathrm{d}t+\sigma_t^S S_t\mathrm{d}W_T^S,\quad S_0=s_0.$$

We assume that $b>0$ and that the processes $\mu^R,\mu^S,\sigma^S:\Omega\times[0,T]\to\mathbb{R}$ are adapted and that $\sigma^S>0$. We write $\sigma_t=(\sigma_t^S S_t,0)^T$.

Lemma 7.19 ([HM07]). *Any linear pricing scheme on the set $L^2(\mathbb{P})$ of square-integrable random variables with respect to $\mathbb{P}$ can be identified with a 2-dimensional predictable process $\theta = (\theta^S, \theta^R)$ such that the exponential process $(\mathcal{E}^\theta_t)$ defined by*

$$\mathcal{E}^\theta_t := \mathcal{E}\left(-\int_0^\cdot \langle \theta_s, \mathrm{d}W_s\rangle\right)_t = \exp\left\{-\int_0^t \langle \theta_s, \mathrm{d}W_s\rangle - \frac{1}{2}\int_0^t |\theta_s|^2 \mathrm{d}s\right\}, \quad t \in [0,T], \quad (7.35)$$

is a uniformly integrable martingale.

This ensures that the measure $\mathbb{P}^\theta$ defined by having density $\mathcal{E}^\theta_T$ against $\mathbb{P}$ is indeed a probability measure (the *pricing measure*), and the present price of a random terminal payment X is given by $\mathbb{E}^\theta[X]$, where $\mathbb{E}^\theta$ denotes the expectation with respect to $\mathbb{P}^\theta$. For any such θ, we introduce the $\mathbb{P}^\theta$-Brownian motion

$$W^\theta_t = W_t + \int_0^t \theta_s \,\mathrm{d}s, \qquad t \in [0,T]. \quad (7.36)$$

The first component θ^S of the vector θ is the *market price of financial risk*. Under the assumption that there is no arbitrage, S must be a martingale under $\mathbb{P}^\theta$ and, from the exogenously given dynamics of S, θ^S is necessarily given by $\theta^S_t = \mu^S_t/\sigma^S_t$, $t \in [0,T]$. The process θ^R on the other hand is unknown. It is the *market price of external risk* and will be derived endogenously by the market clearing condition (or constant net supply condition, see below).

Proof of Lemma 7.19. By the Riesz representation theorem, any continuous positive linear pricing rule $l\colon L^2(\mathbb{P}) \to \mathbb{R}$ can be identified with a square integrable random variable G such that

$$l(F) = \mathbb{E}\left[G \cdot F\right], \quad \forall F \in L^2(\mathbb{P}).$$

If F is a security that pays 1 Dollar in every state of the world, then we must have $\mathbb{E}\left[G\right] = \mathbb{E}\left[G \cdot 1\right] = l(F) = 1$, i.e., G is a $\mathbb{P}$-a.s. strictly positive density function. Hence, every pricing measure can be identified with a probability measure $\mathbb{Q}$ that is equivalent to $\mathbb{P}$. Then, in order to have no arbitrage under that pricing measure, the discounted asset price has to be a $\mathbb{Q}$-martingale. The density $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t} =: Z_t$ is an almost surely strictly positive uniformly integrable martingale and can be written as stochastic exponential, i.e., there exists θ such that

$$Z_t = \exp\left(-\int_0^t \theta_s \cdot \mathrm{d}W_s - \frac{1}{2}\int_0^t |\theta_s|^2 \mathrm{d}s\right).$$

□

The agents $a \in \mathbb{A}$ receive at time T the income H^a which depends on the financial and external risk factors. While the agents are able to trade in the financial market to hedge away some of their financial risk, a basis risk remains, originating in the agents' exposure to the non-tradable risk process R. A derivative with payoff H^D at maturity time T is introduced such that, by trading in the derivative H^D, the agents have now a way to reduce their basis risk.

We make the following assumptions on the endowments, derivative payoff and coefficients appearing in the dynamics of S and R.

Assumption 7.20. *The processes μ^R, μ^S, σ^S and $\theta^S := \mu^S/\sigma^S$ are bounded (belong to $\mathcal{S}^\infty$). The random variables H^D and H^a, $a \in \mathbb{A}$, are bounded (belong to $L^\infty(\mathcal{F}_T)$).*

For a probability measure $\mathbb{Q}$, we denote $\mathcal{H}_{\text{BMO}}(\mathbb{Q})$ as the space of processes $Z \in \mathcal{H}^p(\mathbb{Q})$ for any $p \geq 2$ such that for some constant $K_{BMO} > 0$

$$\sup_{\tau \in \mathcal{T}_{[0,T]}} \mathbb{E}^{\mathbb{Q}} \left[\int_\tau^T |Z_s|^2 \mathrm{d}s \Big| \mathcal{F}_\tau \right] \leq K_{BMO} < \infty,$$

where $\mathcal{T}_{[0,T]}$ is the set of all stopping times $\tau \in [0,T]$. For the reference measure $\mathbb{P}$ we write directly $\mathcal{H}_{\text{BMO}}$ instead of $\mathcal{H}_{\text{BMO}}(\mathbb{P})$.

Assuming no arbitrage opportunities, the price process $(B_t^\theta)_{t\in[0,T]}$ of H^D is given by its expected payoff under $\mathbb{P}^\theta$; in other words $B_\cdot^\theta = \mathbb{E}^\theta \left[H^D | \mathcal{F}_\cdot\right]$. Since H^D is bounded, writing the $\mathbb{P}^\theta$-martingale as a stochastic integral against the $\mathbb{P}^\theta$-Brownian motion W^θ (with the martingale representation theorem) yields a 2-dimensional square-integrable adapted process $\kappa^\theta := (\kappa^S, \kappa^R)$ such that for $t \in [0,T]$

$$B_t^\theta = \mathbb{E}^\theta[H^D] + \int_0^t \langle \kappa_s^\theta, \mathrm{d}W_s^\theta \rangle = \mathbb{E}^\theta[H^D] + \int_0^t \langle \kappa_s^\theta, \mathrm{d}W_s \rangle + \int_0^t \left\langle \kappa_s^\theta, \theta_s \right\rangle \mathrm{d}s. \tag{7.37}$$

We have $(B, \kappa) \in \mathcal{S}^\infty \times \mathcal{H}_{\text{BMO}}(\mathbb{P}^\theta)$. It will turn out to be useful to rewrite (7.37) as a BSDE:

$$B_t^\theta = H^D - \int_t^T \langle \kappa_s^\theta, \theta_s \rangle \mathrm{d}s - \int_t^T \langle \kappa_s^\theta, \mathrm{d}W_s \rangle. \tag{7.38}$$

We denote by $\pi_t^{a,1}$ and $\pi_t^{a,2}$ the number of units agent $a \in \mathbb{A}$ holds in the stock and the derivative at time $t \in [0,T]$, respectively. Using a self-financing strategy $\pi^a := (\pi^{a,1}, \pi^{a,2})$ with values in $\mathbb{R}^2$, her gains from trading up to time $t \in [0,T]$, under the pricing measure $\mathbb{P}^\theta$ inducing the prices (B_t^θ) for the derivative, are given by

$$\begin{aligned} V_t^a = V_t(\pi^a) &= \int_0^t \pi_s^{a,1} \, \mathrm{d}S_s + \int_0^t \pi_s^{a,2} \, \mathrm{d}B_s^\theta \\ &= \int_0^t \left\langle \pi_s^{a,1} \sigma_s + \pi_s^{a,2} \kappa_s^\theta, \theta_s \right\rangle \, \mathrm{d}s + \int_0^t \langle \pi_s^{a,1} \sigma_s + \pi_s^{a,2} \kappa_s^\theta, \mathrm{d}W_s \rangle. \end{aligned} \tag{7.39}$$

We require that the trading strategies be integrable against the prices, i.e., for all $a \in \mathbb{A}$, $\pi^a \in L^2\big((S, B^\theta), \mathbb{P}^\theta\big)$, so that the gains processes are square-integrable martingales under $\mathbb{P}^\theta$ (i.e., we require $\mathbb{E}^\theta \left[\langle \, V_\cdot(\pi^a) \, \rangle_T\right] < \infty$). The $(\mathbb{R}^2)^{(N-1)}$-valued vector of strategies of all agents $b \in \mathbb{A} \setminus \{a\}$ will be denoted by $\pi^{-a} := (\pi^b)_{b \in \mathbb{A}\setminus\{a\}}$.

The agents' measure of risk

The agents assess their risk using a dynamic convex time-consistent risk measure $\rho_\cdot^a$ induced by a BSDE.This means that the risk $\rho_t^a(\xi^a)$ which agent $a \in \mathbb{A}$ associates at time $t \in [0,T]$ with an $\mathcal{F}_T$-measurable random position ξ^a is given by Y_t^a, where (Y^a, Z^a) is the solution to the BSDE

$$-\mathrm{d}Y_t^a = g^a(t, Z_t^a)\mathrm{d}t - \langle Z_t^a, \mathrm{d}W_t \rangle \quad \text{with terminal condition} \quad Y_T^a = -\xi^a.$$

The driver g^a encodes the risk preferences of agent a for $a \in \mathbb{A}$. We assume that g^a has the following properties:

Assumption 7.21. *The map $g^a : [0,T]\times\mathbb{R}^2 \to \mathbb{R}$ is a deterministic continuous function. Its restriction to the space variable, $z \mapsto g^a(\cdot, z)$, is continuously differentiable, strictly convex and attains its minimum.*

For any fixed $(t,\vartheta) \in [0,T] \times \mathbb{R}^2$, the map $z \mapsto g^a(t,z) - \langle z, \vartheta\rangle$ is also strictly convex and attains its unique minimum at the point where its gradient vanishes. With this in mind we can define $\mathcal{Z}^a : [0,T] \times \mathbb{R}^2 \to \mathbb{R}^2$, $(t,\vartheta) \mapsto \mathcal{Z}^a(t,\vartheta)$ where $\mathcal{Z}^a(t,\vartheta)$ is the unique solution, in the unknown $\mathcal{Z}$, to the equation[31]

$$\nabla_z g^a(t,\mathcal{Z}) = \vartheta. \tag{7.40}$$

Under Assumption 7.21, the risk measure given by the above BSDE is strongly time consistent, convex and translation invariant (or monetary).

Agent a's position at maturity, ξ^a, is given by the sum of her terminal income H^a and the trading gains V^a_T over the time period $[0,T]$, i.e.,

$$\xi^a = H^a + V_T(\pi^a).$$

The risk associated with the self-financing strategy π^a evolves according to the BSDE

$$\begin{aligned} -\mathrm{d}Y^a_t &= g^a\left(t, Z^a_t\right)\mathrm{d}t - \langle Z^a_t, \mathrm{d}W_t\rangle, \quad t \in [0,T] \\ Y^a_T &= -\xi^a(\pi^a) = -\Big(H^a + V_T(\pi^a)\Big). \end{aligned} \tag{7.41}$$

Now, we introduce a notion of admissibility for our problem.

Definition 7.22 (Admissibility). *For $a \in \mathbb{A}$, the $\mathbb{R}^2$-valued strategy process π^a is called* admissible *with respect to the market price of risk θ if $\mathbb{E}^\theta\left[\langle\, V_\cdot(\pi^a)\,\rangle_T\right] < \infty$, where $\langle V_\cdot(\pi^a)\rangle$ denotes the quadratic variation of $\big(V_t(\pi^a)\big)_{t\in[0,T]}$, and BSDE* (7.41) *has a unique solution. The set of admissible trading strategies for agent $a \in \mathbb{A}$ is denoted by $\mathcal{A}^\theta$.*

Each agent $a \in \mathbb{A}$ wants to minimize her risk, i.e. agent a solves the risk-minimization problem

$$\min_{\pi^a \in \mathcal{A}^\theta} Y^a_0(\pi^a). \tag{7.42}$$

Equilibrium market price of risk and endogenous trading

We denote by $n \in \mathbb{R}$ the number of units of derivative present in the market. While each unit of derivative pays H^D at time T, the agents are free to buy and underwrite contracts for any amount of H^D. Within the trading period $[0,T]$, only the agents in our set $\mathbb{A}$, with trading objectives as described above, are active in the market and the total number n of derivatives present is constant over time.

We show that one can convert the problem for a general $n \in \mathbb{R}$ into a problem for $n = 0$ by distributing the derivative among all agents before the beginning of the trading period. In the case $n = 0$ every derivative held by an agent has been underwritten by another agent in $\mathbb{A}$, entailing essentially that agents share their risks with each other.

We assume that each agent seeks to minimize her risk measure independently, without cooperation with the other agents, so we are interested in NE.

[31] We write $\nabla_z g^a(t,\mathcal{Z})$ for the vector consisting of the partial derivatives of g^a w.r.t. the (two) components of the space variable, evaluated at $\mathcal{Z}$.

Definition 7.23 (Equilibrium and Equilibrium MPR (EMPR)). *For a given* Market Price of Risk *(MPR)* $\theta = (\theta^S, \theta^R)$, *we call* $\pi^* = (\pi^{*,a})_{a\in\mathbb{A}}$ *an* equilibrium *if, for all* $a \in \mathbb{A}$, $\pi^{*,a} \in \mathcal{A}^\theta$ *and*

for any admissible strategy π^a *it holds that* $Y_0^a(\pi^{*,a}) \leq Y_0^a(\pi^a)$,

i.e., individual optimality. We call θ Equilibrium Market Price of Risk *(EMPR) and* θ^R EMP of external Risk *(EMPeR) if*

1. $\theta = (\theta^S, \theta^R)$ *makes* $\mathbb{P}^\theta$ *a true probability measure (equivalently,* $\mathcal{E}^\theta$ *from* (7.35) *is a uniformly integrable martingale);*
2. *there exists a unique equilibrium* π^* *for* θ*;*
3. π^* *satisfies the market clearing condition (or fixed supply condition) for the derivative* H^D *(where* Leb *denotes the Lebesgue measure):*

$$\sum_{a\in\mathbb{A}} \pi_t^{*,a,2} = \sum_{a\in\mathbb{A}} \pi_0^{*,a,2} = n \quad \mathbb{P}\otimes Leb - a.e.. \tag{7.43}$$

Optimal response for one agent - optimizing the residual risk

In this subsection, in addition to a MPR θ being fixed, we focus on a single agent $a \in \mathbb{A}$, whose preferences are encoded by g^a, we take the strategies $\pi^{-a} = (\pi^b)_{b\in\mathbb{A}\setminus\{a\}}$ of the other agents as given, and we study the investment problem of our agent in this setting.

To solve the optimization problem (7.42) for agent a, we first recall from [HPDR10] that, at each time $t \in [0,T]$, the strategy chosen must minimize the residual risk: translation invariance implies that

$$Y_t^a = \rho_t^a\Big(H^a + V_T^a\Big) = \rho_t^a\Big(H^a + (V_T^a - V_t^a)\Big) - V_t^a.$$

This suggests applying the following change of variables to (7.41),

$$\begin{aligned} \widetilde{Y}_t^a &:= Y_t^a + V_t^a, \\ \widetilde{Z}_t^a &:= Z_t^a + \zeta_t^a, \quad \text{where} \quad \zeta_t^a = \pi_t^{a,1}\sigma_t + \pi_t^{a,2}\kappa_t^\theta \in \mathbb{R}^2. \end{aligned} \tag{7.44}$$

If the strategies are not clear from the context, we also write $\zeta^a = \zeta^a(\pi^a)$. Direct computations yield a BSDE for $(\widetilde{Y}^a, \widetilde{Z}^a)$ given by

$$\begin{aligned} -d\widetilde{Y}_t^a &= \widetilde{g}^a\Big(t, \pi_t^a, \widetilde{Z}_t^a\Big)\mathrm{d}t - \langle \widetilde{Z}_t^a, \mathrm{d}W_t\rangle, \quad t \in [0,T], \\ \widetilde{Y}_T^a &= -H^a, \end{aligned} \tag{7.45}$$

where the driver $\widetilde{g}^a\colon \Omega \times [0,T] \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$\widetilde{g}^a(t, \pi_t^a, z^a) := g^a\big(t, z^a - \zeta_t^a\big) - \langle \zeta_t^a, \theta_t\rangle \tag{7.46}$$

$$= g^a\bigg(t, z^a - \Big(\pi_t^{a,1}\sigma_t + \pi_t^{a,2}\kappa_t^\theta\Big)\bigg) - \Big\langle \pi_t^{a,1}\sigma_t + \pi_t^{a,2}\kappa_t^\theta, \theta_t\Big\rangle. \tag{7.47}$$

Each individual agent $a \in \mathbb{A}$ seeks to minimize $\widetilde{Y}_0^a$, the solution to (7.45), via her choice of investment strategy $\pi^a \in \mathcal{A}^\theta$, in other words she aims at solving

$$\min_{\pi^a \in \mathcal{A}^\theta} \widetilde{Y}_0^a(\pi^a). \tag{7.48}$$

Before we solve the individual optimization problem, we make the assumption that the derivative H^D does indeed complete the market. This must then be verified a posteriori (once the solution is computed) and case-by-case depending on the specific model.

Assumption 7.24. *Assume that $\kappa_t^R \neq 0$, for any $t \in [0,T]$, $\mathbb{P}$-a.s.* .

The pointwise minimizer for the single agent's residual risk

In (7.45), the strategy π^a appears only in the driver $\widetilde{g}^a$, not in the terminal condition. The comparison theorem for BSDEs suggests that in order to minimize $\widetilde{Y}_0^a(\pi^a)$ over admissible strategies π^a, one needs only to minimize the driver function $\widetilde{g}^a$ over π_t^a, for each fixed ω, t and z^a. We define such pointwise minimizer as the random map $\Pi^a \colon \Omega \times [0,T] \times \mathbb{R}^2 \to \mathbb{R}^2$, given by

$$\Pi^a(t,z) := \arg\min_{\pi^a \in \mathbb{R}^2} \widetilde{g}^a(t,\pi^a,z).$$

The pointwise minimization problem has, under Assumption 7.21, a unique minimizer, which is characterized by the FOC for $\widetilde{g}^a$, i.e.,

$$\nabla_{\pi^a} \widetilde{g}^a(t,\pi^a,z^a) = 0.$$

Recall that $\sigma = (\sigma^S S, 0)$. Using (7.47), the FOC is equivalently written as

$$\begin{aligned}
\partial_{\pi^{a,1}} \widetilde{g}^a(t,\pi^a,z^a) = 0 &\iff \big\langle (\nabla_z g^a)(t, z^a - \zeta^a), -\sigma \big\rangle - \langle \sigma, \theta \rangle = 0 \\
&\iff g^a_{z^1}(t, z^a - \zeta^a) = -\theta^S, \qquad\qquad (7.49)\\
\partial_{\pi^{a,2}} \widetilde{g}^a(t,\pi^a,z^a) = 0 &\iff \big\langle (\nabla_z g^a)(t, z^a - \zeta^a), -\kappa^\theta \big\rangle - \langle \kappa^\theta, \theta \rangle = 0 \\
&\iff -\theta^S \kappa^S + g^a_{z^2}(t, z^a - \zeta^a)\, \kappa^R = -\kappa^S \theta^S - \kappa^R \theta^R \\
&\iff g^a_{z^2}(t, z^a - \zeta^a) = -\theta^R, \qquad\qquad (7.50)
\end{aligned}$$

where we used (7.49) to obtain (7.50) under Assumption 7.24.

With $\mathcal{Z}^a$ from (7.40), the FOC system (7.49)–(7.50) is equivalent to

$$z^a - \zeta_t^a = z^a - \zeta_t^a(\Pi^a(t,z^a)) = \mathcal{Z}^a(t, -\theta_t), \tag{7.51}$$

which has the useful property that while the LHS depends on z^a, the RHS merely depends on the MPR θ and the structure of the driver g^a.

The expression for ζ^a in (7.44) and elementary re-arrangements allow to rewrite (7.51) as

$$\begin{aligned}
\Pi^{a,1}(t,z^a) &= \frac{z^{a,1} - \mathcal{Z}^{a,1}(t,-\theta_t)}{\sigma_t^S S_t} - \frac{z^{a,2} - \mathcal{Z}^{a,2}(t,-\theta_t)}{\kappa_t^R} \, \frac{\kappa_t^S}{\sigma_t^S S_t}, \\
\Pi^{a,2}(t,z^a) &= \frac{z^{a,2} - \mathcal{Z}^{a,2}(t,-\theta_t)}{\kappa_t^R}.
\end{aligned} \tag{7.52}$$

Plugging $z^a - \zeta_t^a = \mathcal{Z}^a(t, -\theta_t)$ into (7.46) yields an expression for the minimized (random) driver

$$\begin{aligned}\widetilde{g}^a\big(t, \Pi^a(t, \pi_t^{-a}, z^a), z^a\big) &= g^a\big(t, \mathcal{Z}^a(t, -\theta_t)\big) + \langle \mathcal{Z}^a(t, -\theta_t), \theta_t\rangle - \langle z^a, \theta_t\rangle \\ &=: \widetilde{G}^a(t, z^a).\end{aligned} \tag{7.53}$$

Observe that $\widetilde{G}^a$ is an affine driver (in z^a) with stochastic coefficients.

Single-agent optimality

Since $\widetilde{G}^a$ is an affine driver, and since $\nabla_z \widetilde{G}^a = -\theta \in \mathcal{H}_{\text{BMO}}$, we have a unique solution to the BSDE with driver $\widetilde{G}^a$ and terminal condition $-H^a$, provided that the process $(\omega, t) \mapsto \widetilde{G}^a(t, 0) = g^a\big(t, \mathcal{Z}^a(t, -\theta_t)\big) + \langle \mathcal{Z}^a(t, -\theta_t), \theta_t\rangle$ is sufficiently integrable. Let $(\widetilde{Y}^a, \widetilde{Z}^a)$ be the solution to BSDE (7.45) with driver (7.53) and define the strategy $\pi^{*,a}_\cdot := \Pi^a(\cdot, \widetilde{Z}^a_\cdot)$. This does not only solve the individual risk minimization problem, but, as we show next, it is even unique.

Theorem 7.25 (Optimality for one agent, uniqueness[32]). *Fix a market price of risk $\theta = (\theta^S, \theta^R) \in \mathcal{H}_{BMO}$ and let Assumption 7.24 hold. Fix an agent $a \in \mathbb{A}$ and a set of integrable strategies $(\pi^b)_{b \in \mathbb{A}\setminus\{a\}}$. Assume further that for $\widetilde{G}^a$ given by (7.53), $|\widetilde{G}^a(\cdot, 0)|^{1/2} \in \mathcal{H}_{BMO}$. Then the BSDE with driver (7.53) and terminal condition $-H^a$ has a unique solution $(\widetilde{Y}^a, \widetilde{Z}^a) \in \mathcal{S}^\infty \times \mathcal{H}_{BMO}$. Moreover, if $\pi^{*,a}_\cdot = \Pi^a(\cdot, \widetilde{Z}^a_\cdot)$ is admissible, then $\widetilde{Y}_0^a$ is the value of the optimization problem (7.48) (i.e. the minimized risk) for agent a and $\pi^{*,a}$ is the unique optimal strategy.*

Proof. Given the structure of $\widetilde{G}^a$ in (7.53) and the integrability assumption made, the existence and uniqueness of the BSDE's solution $(\widetilde{Y}^a, \widetilde{Z}^a)$ in $\mathcal{S}^\infty \times \mathcal{H}_{\text{BMO}}$ is straightforward[33].

We first use the comparison theorem to prove the minimality of $\widetilde{Y}^a$, and hence the optimality of $\pi^{*,a}$. Let $t \in [0, T]$. Take any strategy $\pi^a \in \mathcal{A}^\theta$. First, from the definition of $\widetilde{G}^a$ as a pointwise minimum, we naturally have that

$$\widetilde{G}^a(t, z^a) = \widetilde{g}^a\left(t, \Pi^a(t, z^a), z^a\right) \le \widetilde{g}^a(t, \pi_t^a, z^a) \text{ for all } t \in [0, T] \text{ and } z^a \in \mathbb{R}^2,$$

i.e., $\widetilde{G}^a(\cdot, \cdot) \le \widetilde{g}^a(\cdot, \pi^a_\cdot, \cdot)$. Second, $\widetilde{G}^a$ is affine and thus Lipschitz continuous, with Lipschitz coefficient process $-\theta \in \mathcal{H}_{\text{BMO}}$. By the comparison theorem, we therefore have, for any $t \in [0, T]$ and in particular for $t = 0$, that $\widetilde{Y}_t^a = \widetilde{Y}_t^a(\pi^{*,a}) \le \widetilde{Y}_t^a(\pi^a)$. As this holds for any $\pi^a \in \mathcal{A}^\theta$, this proves the minimality of $\widetilde{Y}_0^a = \rho_0^a\big(\xi^a(\pi^{*,a})\big)$ and thus the optimality of $\pi^{*,a}$.

We now argue the uniqueness of the optimizer $\pi^{*,a}$. Let π^a be an admissible strategy and let $(\widetilde{Y}^a(\pi^a), \widetilde{Z}^a(\pi^a))$ be the corresponding risk, i.e. solution to the BSDE (7.45) with

[32] cf. [HPDR10]*Proposition 3.6, which we extend by proving existence and uniqueness of the solution of the BSDE instead of assuming it. Uniqueness is important to us in order to obtain a unique NE for a given MPR θ.

[33] cf. [IDR10]*Theorem 2.6, which states that $Y \in \mathcal{S}^\infty$ and $Z * W \in BMO$, which implies $Z \in \mathcal{H}_{\text{BMO}}$.

strategy π^a. We compute the difference $\widetilde{Y}^a_t(\pi^a) - \widetilde{Y}^a_t(\pi^{*,a})$:

$$
\begin{aligned}
&\widetilde{Y}^a_t(\pi^a) - \widetilde{Y}^a_t(\pi^{*,a}) \\
&= \int_t^T \left[\widetilde{g}^a\left(s, \pi^a_s, \widetilde{Z}^a_s(\pi^a)\right) - \widetilde{G}^a\left(s, \widetilde{Z}^a_s(\pi^{*,a})\right)\right] \mathrm{d}s - \int_t^T [\widetilde{Z}^a_s(\pi^a) - \widetilde{Z}^a_s(\pi^{*,a})]\mathrm{d}W_s \\
&= \int_t^T \left[\widetilde{g}^a\left(s, \pi^a_s, \widetilde{Z}^a_s(\pi^a)\right) - \widetilde{g}^a\left(s, \Pi^a\big(s, \widetilde{Z}^a_s(\pi^a)\big), \widetilde{Z}^a_s(\pi^a)\right)\right] \mathrm{d}s \qquad (7.54) \\
&\quad - \int_t^T [\widetilde{Z}^a_s(\pi^a) - \widetilde{Z}^a_s(\pi^{*,a})]\mathrm{d}W^\theta_s,
\end{aligned}
$$

where we added $\widetilde{G}^a(t, \widetilde{Z}^a_t(\pi^a))$, subtracted $\widetilde{g}^a\left(t, \Pi^a\big(t, \widetilde{Z}^a_t(\pi^a)\big), \widetilde{Z}^a_t(\pi^a)\right)$ (equal to the added term) and used the affine structure[34] of $\widetilde{G}^a$ combined with (7.36).
By construction of Π^a as a minimizer, the difference in (7.54) is always positive. In particular, taking $\mathbb{P}^\theta$-expectation w.r.t. $\mathcal{F}_t$ implies that $\widetilde{Y}^a_t(\pi^a) - \widetilde{Y}^a_t(\pi^{*,a}) \geq 0$ for all $t \in [0, T]$. Assume that π^a is an optimal strategy. Then $\widetilde{Y}^a_0(\pi^a) = \widetilde{Y}^a_0(\pi^{*,a})$ and the LHS vanishes for $t = 0$. Under $\mathbb{P}^\theta$-expectation, the stochastic integral on the RHS also vanishes and we can conclude that the integrand in (7.54) is zero $\mathbb{P}^\theta \otimes Leb$-a.e.. Consequently, we obtain $\widetilde{Y}^a(\pi^a) = \widetilde{Y}^a(\pi^{*,a})$ and hence $\widetilde{Z}^a(\pi^a) = \widetilde{Z}^a(\pi^{*,a})$. This implies $\Pi^a\big(\cdot, \widetilde{Z}^a_\cdot(\pi^a)\big) = \Pi^a\big(\cdot, \widetilde{Z}^a_\cdot(\pi^{*,a})\big)$. Finally, by uniqueness of the minimizer Π^a, we obtain $\pi^a_\cdot = \Pi^a\big(\cdot, \widetilde{Z}^a_\cdot(\pi^a)\big) = \Pi^a\big(\cdot, \widetilde{Z}^a_\cdot(\pi^{*,a})\big) = \pi^{*,a}_\cdot$. $\square$

Remark 7.26. *Theorem 7.25 describes the optimal investment of an agent with preferences described by g^a (equivalently, $\rho^a_\cdot$) who trades in the assets S and B, which have the given MPR θ. Following the same methods, the result could be generalized to a higher number of assets, with price processes given exogenously. This applies similarly to an agent trading in fewer assets, by setting the respective components to zero.*

We now state a characterization of the optimal strategy via the FOC.

Lemma 7.27. *Under the assumptions of Theorem 7.25, let $\widehat{\pi}^a$ be an admissible strategy and $(\widehat{Y}^a, \widehat{Z}^a)$ be the associated risk process, solution to the BSDE with driver $\widetilde{g}^a(t, \widehat{\pi}^a_t, \pi^{-a}_t, \cdot)$ and terminal condition $-H^a$. Assume that they satisfy the FOC* (7.49)–(7.50) *in the sense that*

$$
\nabla_z g^a(t, \widehat{Z}^a_t - \widehat{\zeta}^a_t) = -\theta_t \qquad \textit{where} \qquad \widehat{\zeta}^a_t = \left(\widehat{\pi}^{a,1}_t \sigma_t + \widehat{\pi}^{a,2}_t \kappa^\theta_t\right) - \widetilde{\lambda}^a\left(\bar{\pi}^{-a,1}_t \sigma_t + \bar{\pi}^{-a,2}_t \kappa^\theta_t\right).
$$

Then $(\widehat{Y}^a, \widehat{Z}^a) = (\widetilde{Y}^a, \widetilde{Z}^a)$ and $\widehat{\pi}^a = \pi^{*,a}$.

Proof. By the assumptions on g^a, $\nabla_z g^a(t, \widehat{Z}^a_t - \widehat{\zeta}^a_t) = -\theta_t$ means that $\widehat{Z}^a_t - \widehat{\zeta}^a_t = \mathcal{Z}^a(t, -\theta_t)$, or equivalently $\widehat{\pi}_t = \Pi^a(t, \widehat{Z}^a_t)$. Therefore, $\widetilde{g}^a(t, \widehat{\pi}^a_t, \widehat{Z}^a_t) = \widetilde{G}^a(t, \widehat{Z}^a_t)$ – recall (7.46). By uniqueness of the solution to the BSDE with driver $\widetilde{G}^a(t, \cdot)$ and terminal condition $-H^a$, we have $(\widehat{Y}^a, \widehat{Z}^a) = (\widetilde{Y}^a, \widetilde{Z}^a)$. Consequently, by the uniqueness of the FOC's solution (Theorem 7.25), $\widehat{\pi}^a_t = \Pi^a(t, \widehat{Z}^a_t) = \Pi^a(t, \widetilde{Z}^a_t) = \pi^{*,a}_t$. $\square$

The final step is to calculate the EMPeR θ^R from the market clearing condition. As the optimal strategies π^a are very abstract, let us look at this step in the special setting of the entropic risk measure.

[34] (7.53) implies $\widetilde{G}^a(t, z^a) - \widetilde{G}^a(t, \widehat{z}^a) = \langle \widehat{z}^a - z^a, \theta_t \rangle$ for any $t \in [0, T]$ and $z^a, \widehat{z}^a \in \mathbb{R}^2$.

Example 7.28 (Entropic risk measure)**.** *Each agent $a \in \mathbb{A}$ is assessing her risk*[35] *using the entropic risk measure ρ_0^a for which the driver $g^a : \mathbb{R}^2 \to \mathbb{R}$ is given by*

$$g^a(z) := \frac{1}{2\gamma_a}|z|^2, \quad \textit{where} \quad \gamma_a > 0 \quad \textit{is agent } a\textit{'s} \text{ risk tolerance}, \tag{7.55}$$

and $1/\gamma_a$ is agent a's risk aversion. *Since $g^a_{z^i}(z) = z^i/\gamma_a$, it is easily found that $\mathcal{Z}^a(t, -\theta_t) = (-\gamma_a\theta_t^S, -\gamma_a\theta_t^R) = -\gamma_a\theta_t$ for all $a \in \mathbb{A}$ and $i \in \{1, 2\}$ (cf.* (7.40)*). Injecting this in* (7.53) *yields the minimized driver $\widetilde{G}^a$,*

$$\widetilde{G}^a(t, z^a) = -\frac{\gamma_a}{2}\,|\theta_t|^2 - \langle z^a, \theta_t \rangle\,, \quad t \in [0, T]. \tag{7.56}$$

The minimized (individual) risk is then given by $Y_0^a = \widetilde{Y}_0^a$ where $(\widetilde{Y}^a, \widetilde{Z}^a)$ is the solution to the BSDE with terminal condition $-H^a$ and driver $\widetilde{G}^a$, while the optimal strategies $\pi^* = (\pi^{*,a})_{a\in\mathbb{A}}$ are given by (cf. (7.52)*)*

$$\pi^{*,a,1} = \frac{\widetilde{Z}^{a,1} + \gamma_a\theta^S}{\sigma^S S} - \frac{\widetilde{Z}^{a,2} + \gamma_a\theta^R}{\kappa^R}\,\frac{\kappa^S}{\sigma^S S}, \tag{7.57}$$

$$\pi^{*,a,2} = \frac{\widetilde{Z}^{a,2} + \gamma_a\theta^R}{\kappa^R}. \tag{7.58}$$

If one assumes $n = 0$, then the market clearing condition (7.43) *reads $\sum_{a\in\mathbb{A}} \pi^{a,2} = 0$. With this, the market price of external risk θ^R can be computed from adding* (7.58) *over $a \in \mathbb{A}$, giving $\theta^R = -\sum_{a\in\mathbb{A}} \widetilde{Z}^{a,2} / \sum_{a\in\mathbb{A}} \gamma_a$.*

[35] Our problem is that of minimizing a static risk measure. The corresponding dynamic risk measure is be given by $\rho_t^a(\xi) := \gamma_a \ln \mathbb{E}[e^{-\xi/\gamma_a} | \mathcal{F}_t]$ for $t \in [0, T]$.

A Collection of results for the stochastic exponential

The stochastic exponential of an Itô process L with $L_0 = 0$ is

$$\mathcal{E}(L)_t = \exp\left(L_t - \frac{1}{2}\langle L\rangle_t\right).$$

$U = \mathcal{E}(L)$ satisfies the SDE

$$\mathrm{d}U_t = U_t \mathrm{d}L_t, \quad \text{with initial condition} \quad U_0 = 1,$$

hence it has the integral representation

$$U_t = 1 + \int_0^t U_s \mathrm{d}L_s.$$

Lemma A.1. *If L is a local martingale and H locally bounded, then $\int_0^\cdot H_t \mathrm{d}L_t$ is a local martingale as well.*

Proof. This is an object of Stochastic Analysis. The result and its proof can be found, e.g., as Theorem 29 in Chapter IV in *Stochastic Integration and Differential Equations* by P. E. Protter (2005). □

Lemma A.2. *If L is a local martingale, then $U = \mathcal{E}(L)$ is a local martingale as well.*

Proof. Due to its integral representation, all we need to show is that $\int_0^t U_s \mathrm{d}L_s$ is a local martingale. This follows directly from Lemma A.1. □

Corollary A.3. *The expected value of a stochastic exponential is bounded by 1. In other words, if L is a local martingale with $L_0 = 0$, then $\mathbb{E}\left[\mathcal{E}(L)_\tau\right] \leq 1$ for any finite stopping time τ.*

Proof. The above lemma tells us that $\mathcal{E}(L)$ is a local martingale. It is bounded from below (by zero), hence it is a supermartingale by Lemma 2.20 with $\mathbb{E}\left[\mathcal{E}(L)_0\right] = 1$. Consequently,

$$\mathbb{E}\left[\mathcal{E}(L)_\tau\right] = \mathbb{E}\left[\mathcal{E}(L)_\tau \mid \mathcal{F}_0\right] \leq \mathcal{E}(L)_0 = 1.$$

□

Lemma A.4. *Let S be an Itô process satisfying $\mathrm{d}S_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t$. Assume that in the market there exists an ELMM. Then $\mathbb{Q} = Z_T \mathbb{P}$ for $Z_t = \mathcal{E}(-\int_0^\cdot \xi_s \mathrm{d}W_s)_t$ and with $W_t^* = \int_0^t \xi_s \mathrm{d}s + W_t$ we have $\mathrm{d}S_t = \sigma_t \mathrm{d}W_t^*$.*

Proof. Assume that $\mathbb{P} \approx \mathbb{Q}$. Therefore, a strictly positive Radon-Nikodym derivative Z exists such that $Z_t := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t} > 0$ is a uniformly integrable $\mathbb{P}$-martingale. From the MRT we know that there exists $\eta \in \mathcal{V}$ such that $Z_t = \mathbb{E}^{\mathbb{P}}\left[Z_0\right] + \int_0^t \eta_s \mathrm{d}W_s$. As $Z_t > 0$ we can introduce $\xi_t := -\frac{\eta_t}{Z_t}$ and write

$$Z_t = \mathbb{E}^{\mathbb{P}}\left[Z_0\right] + \int_0^t \frac{\eta_s}{Z_s} Z_s \mathrm{d}W_s = \mathbb{E}^{\mathbb{P}}\left[Z_0\right] + \int_0^t -\xi_s Z_s \mathrm{d}W_s.$$

As (Z_t) satisfies the SDE $\mathrm{d}Z_t = -\xi_t Z_t \mathrm{d}W_t$, we have $Z_t = \mathcal{E}(-\int_0^{\cdot} \xi_s \mathrm{d}W_s)_t$. If we let $W_t^* := \int_0^t \xi_s \mathrm{d}s + W_t$, then Girsanov's first theorem tells us that W^* is a BM w.r.t. $\mathbb{Q}$ and

$$\mathrm{d}S_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t = \mu_t \mathrm{d}t + \sigma_t(\mathrm{d}W_t^* - \xi_t \mathrm{d}t) = (\mu_t - \sigma_t \xi_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t^*.$$

Now S is a continuous local martingale (by assumption on $\mathbb{Q}$ being an ELMM) and $\int_0^t \sigma_s \mathrm{d}W_s^*$ is also a local martingale for $\sigma \in \mathcal{W}$. Consequently, $\int_0^t (\mu_s - \sigma_s \xi_s)\mathrm{d}s$ must be a local martingale as well. It is a local martingale with bounded variation on $[0, T]$. This implies that it is constant and from its initial value being zero we can infer that $\int_0^t (\mu_s - \sigma_s \xi_s)\mathrm{d}s \equiv 0$ almost surely. □

Exercise 13. *Let X be a continuous local martingale with $\langle X \rangle \equiv 0$. Show that X is almost surely constant.*

B Addendum to Section 5

In Section 5 on stochastic optimal control we have considered the restriction to constant controls, i.e., instead of optimizing over $\alpha \in \mathcal{A}$, we only maximized over $a \in A$. For this to make sense, we have to ensure that constant controls are admissible.

While we have made assumptions on g (Assumption 5.3), we have neglected f. The following remark shall fill this gap:

Remark B.1. *Assume that f satisfies a quadratic growth condition, i.e., that there exists a constant $C > 0$ and a positive function $\kappa \colon A \to \mathbb{R}_+$ such that*

$$|f(t, x, a)| \leq C(1 + |x|^2) + \kappa(a), \quad \forall (t, x, a) \in [0, T] \times \mathbb{R}^d \times A.$$

Then, from (5.4) *follows that for all $(t, x) \in [0, T] \times \mathbb{R}^d$ and all constant controls $a \in A$,*

$$\mathbb{E}\left[\int_t^T \left|f(s, X_s^{t,x}, a)\right| \mathrm{d}s\right] < \infty,$$

which entails that $a \in \mathcal{A}(t, x)$, i.e., all constant controls are admissible.

If, moreover, there exists a constant $C' > 0$ such that

$$\kappa(a) \leq C'(1 + |\gamma(0, a)|^2 + |\sigma(0, a)|^2), \quad \forall a \in A,$$

then from the above and the definition of $\mathcal{A}$ follows that for all $(t, x) \in [0, T] \times \mathbb{R}^d$ and for all $\alpha \in \mathcal{A}$,

$$\mathbb{E}\left[\int_t^T \left|f(s, X_s^{t,x}, \alpha_s)\right| \mathrm{d}s\right] < \infty.$$

Hence, in this case we have $\mathcal{A}(t, x) = \mathcal{A}$.

C Multidimensional Itô processes and the multidimensional Itô formula

This is a collection of multidimensional results, among others, from Section 2.2.

let $W = (W^1, \dots, W^m)$ denote an m-dimensional BM.
Let each component of $A = (a^i)_{i=1,\dots,n}$ and $B = (b^{ij})_{i=1,\dots,n;j=1,\dots,m}$ satisfy the conditions we previously imposed on a and b, respectively. Then an n-dimensional process $X = (X^1, \dots, X^n)$ is an Itô process if we have

$$dX_t = A_t \mathrm{d}t + B_t \mathrm{d}W_t,$$

in other words we have

$$\begin{aligned}
\mathrm{d}X_t^1 &= a_t^1 \mathrm{d}t + b_t^{11}\mathrm{d}W_t^1 + \ldots + b_t^{1m}\mathrm{d}W_t^m,\\
&\vdots\\
\mathrm{d}X_t^n &= a_t^n \mathrm{d}t + b_t^{n1}\mathrm{d}W_t^1 + \ldots + b_t^{nm}\mathrm{d}W_t^m.
\end{aligned}$$

Theorem. *(cf. Thm. 2.32) Let X be an Itô process and let $f(t,x) = (f_1(t,x), \dots, f_r(t,x))$ be a C^1-function in time and a C^2-function in space. Then $Y_t(\omega) := f(t, X_t(\omega))$ defines an Itô process with representation*

$$\mathrm{d}Y_t^k = \frac{\partial}{\partial t} f_k(t,X_t)\mathrm{d}t + \sum_i \frac{\partial}{\partial x_i} f_k(t,X_t)\mathrm{d}X_t^i + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j} f_k(t,X_t)\mathrm{d}X_t^i \mathrm{d}X_t^j. \tag{C.1}$$

for $k \in \{1, \dots, r\}$ and the sums running over $i, j \in \{1, \dots, n\}$.

If we plug in the expressions for $\mathrm{d}X^k$ in (C.1), then we obtain

$$\begin{aligned}
\mathrm{d}Y_t^k = &\left(\frac{\partial}{\partial t} f_k(t,X_t) + \sum_i \frac{\partial}{\partial x_i} f_k(t,X_t) a_t^i\right)\mathrm{d}t + \sum_{i,j}\frac{\partial}{\partial x_i} f_k(t,X_t) b_t^{ij}\mathrm{d}W_t^j\\
&+ \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j} f_k(t,X_t)\left(\sum_l b_t^{il} b_t^{jl}\right)\mathrm{d}t
\end{aligned} \tag{C.2}$$

Now we have $\sum_l b_t^{il} b_t^{jl} = (BB^T)_{ij}$. Furthermore, recall from linear algebra that the trace of the product of two matrices $\mathbb{A}$ and $\mathbb{B}$ can be written as

$$\mathrm{tr}(\mathbb{A}^T\mathbb{B}) = \mathrm{tr}(\mathbb{B}^T\mathbb{A}) = \sum_{i,j}\mathbb{A}^{i,j}\mathbb{B}^{i,j}.$$

Hence, by letting $\mathbb{A} = D_x^2 f_k(t,X_t)$ and $\mathbb{B} = (BB^T)_t$, we can rewrite (C.2) as

$$\begin{aligned}
\mathrm{d}Y_t^k = &\left(\frac{\partial}{\partial t} f_k(t,X_t) + D_x f_k(t,X_t)\cdot A_t + \frac{1}{2}\mathrm{tr}\left((BB^T)_t D_x^2 f_k(t,X_t)\right)\right)\mathrm{d}t\\
&+ D_x f_k(t,X_t)\cdot B_t \mathrm{d}W_t.
\end{aligned}$$

References

[BE05] Pauline Barrieu and Nicole El Karoui. Inf-convolution of risk measures and optimal risk transfer. *Finance Stoch.*, 9(2):269–298, 2005.

[BE09] Pauline Barrieu and Nicole El Karoui. Pricing, Hedging, and Designing Derivatives with Risk Measures. In René Carmona, editor, *In R. Carmona (ed.), Indifference Pricing: Theory and Applications.*, PRINCETON SERIES IN FINANCIAL ENGINEERING, chapter 3, pages 77–146. Princeton, NJ: Princeton University Press, 2009.

[Bel03] Richard Bellman. *Dynamic programming*. Mineola, N.Y., dover ed., repr. of the 6. printing 1972 edition, 2003.

[Bel06] Denis R. Bell. *The Malliavin Calculus*. Dover Publications, Mineola, New York, 2006.

[BLDR17] Jana Bielagk, Arnaud Lionnet, and Gonçalo Dos Reis. Equilibrium pricing under relative performance concerns. *arXiv preprint arXiv:1511.04218*, 2017. Accepted by SIAM Journal on Financial Mathematics.

[BS79] Dimitri P. Bertsekas and Steven E. Shreve. *Stochastic optimal control the discrete time case*, volume 139 of *Mathematics in Science and Engineering*. Elsevier Science, Burlington, 1979.

[Car09] René Carmona, editor. *Indifference Pricing: Theory and Applications*. Princeton series in financial engineering. Princeton [u.a.], 2009.

[CF13] Rama Cont and David-Antoine Fournié. Functional Itô calculus and stochastic integral representation of martingales. *The Annals of Probability*, 41(1):109–133, 2013.

[CIL92] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67, 1992.

[Dup09] Bruno Dupire. Functional Itô Calculus. *Bloomberg Portfolio Research paper*, (2009-04), 2009.

[FS04] H. Föllmer and A. Schied. *Stochastic Finance. An Introduction in Discrete Time.* de Gruyter Studies in Mathematics, Germany, 2004.

[Gia06] Emanuela Rosazza Gianin. Risk measures via g-expectations. *Insurance: Mathematics and Economics*, 39(1):19–34, 2006.

[HM07] Ulrich Horst and Matthias Müller. On the Spanning Property of Risk Bonds Priced by Equilibrium. *Mathematics of Operations Research*, 32(4):784–807, 2007.

[HPDR10] Ulrich Horst, Traian Pirvu, and Gonçalo Dos Reis. On securitization, market completion and equilibrium risk transfer. *Mathematics and Financial Economics*, 2(4):211–252, 2010.

[IDR10] Peter Imkeller and Gonçalo Dos Reis. Path regularity and explicit convergence rate for BSDE with truncated quadratic growth. *Stochastic Process. Appl.*, 120(3):348–379, 2010.

[Imk08] Peter Imkeller. *Malliavin's calculus and applications in stochastic control and finance*, volume 1 of *IMPAN Lecture Notes*. Institute of Mathematics, Polish Academy of Sciences, Warsaw, 2008.

[Kle07] Fima C. Klebaner. *Introduction to stochastic calculus with applications*. Imperial College Press, London [u.a.], 2. reprint. edition, 2007.

[Kle08] Achim Klenke. *Wahrscheinlichkeitstheorie*. Berlin, Heidelberg, 2. edition, 2008.

[KS88] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*. Graduate texts in mathematics 113. New York [u.a.], 1. print. edition, 1988.

[Kus75] H. J. Kushner. Existence results for optimal stochastic controls. *Journal of Optimization Theory and Applications*, 15(4):347–359, Apr 1975.

[Kus14] Harold J. Kushner. A partial history of the early development of continuous-time nonlinear stochastic systems theory. *Automatica*, 50(2):303–334, 2014.

[Øk03] Bernt K. Øksendal. *Stochastic differential equations : an introduction with applications*. Universitext. Springer, Berlin [u.a.], 6. edition, 2003.

[Pha09] Huyên Pham. *Continuous-time stochastic control and optimization with financial applications*. Stochastic modelling and applied probability 61. Berlin [u.a.], 2009.

[PR15] Irina Penner and Anthony Réveillac. Risk measures for processes and bsdes. *Finance and Stochastics*, 19(1):23–66, Jan 2015.

[RY99] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*. Grundlehren der mathematischen Wissenschaften, A Series of Comprehensive Studies in Mathematics, 293. Berlin, Heidelberg, third edition, 1999.

[SKY01] J. Michael Steele, I. Karatzas, and M. Yor. *Stochastic Calculus and Financial Applications*. Springer, New York, 2001.