Angewandte Finanzmathematik 2025: Introduction to the Black–Scholes World

Ari-Pekka Perkkiö

July 15, 2025

Contents

1 Practicalities and background

Upon passing the exam, attending and solving the exercises give a bonus to the final grade.

We assume that the following concepts are familiar:

1. Probability space, random variables, expectation, convergence concepts.
2. Conditional expectations, martingales.
3. The fundamentals of discrete time financial mathematics.

For a remote graphical access to Matlab, you can login to the computers

- math12.math.lmu.de
- mathw0g.math.lmu.de

You will need

1. a program supporting X11-forwarding (e.g. Cygwin),
2. SSH program with rdp connections (e.g. Bitvise),
3. a VPN connection to LRZ (Anyconnect client, downloadable from LMU service portal).

Alternatively, you can use the online version of Matlab.

2 Introduction

2.1 Popular financial products

Throughout the course, S_t^i denotes the market price of an asset i at time t.

Example 2.1 (Put and call options)**.** *A* European call option *on the asset i is a contract where the seller has the obligation to deliver the asset i at the given* maturity time *T for a given* strike price *K. At time T, the buyer has the possibility to exercise the option, that is, to buy the asset from the seller at price K. The gain for the buyer is*

$$c_C := (S_T^i - K)^+ := \max\{S_T^i - K, 0\},$$

since he can get the asset from the option seller at price K and sell it immediately on the market with the market price S_T^i. We call c_C the payoff *of the call option.*

A European put option *on the asset i is a contract where the seller has the obligation to buy the asset i at the given* maturity time T *for a given* strike price K. *At time T, the buyer has the possibility the exercise the option, that is, to sell the asset to the seller at price K. The payoff for the buyer becomes*

$$c_P := (K - S_T^i)^+ := \max\{K - S_T^i, 0\}.$$

Put and call options are prototype examples of *Vanilla options* that depend only on the terminal price of the underlying asset. When this is not the case, the option is called *path-dependent*.

Example 2.2 (Asian options)**.** *An* Asian call option *with maturity T and strike K has the payoff*

$$c_{AC} := (\bar{S}_T^i - K)^+,$$

where $\bar{S}_T^i$ is the "average price" of the asset over the time interval $[0, T]$. The exact form of the average price is part of the contract, e.g., it could be arithmetic mean of the prices at given time points $t_1, \dots, t_N = T$ so that $\bar{S}_T^i = \frac{1}{N}\sum_{k=1}^N S_{t_k}^i$.

For a set A, we denote $\mathbb{1}_A(s) = 1$ if $s \in A$ and $\mathbb{1}_A(s) = 0$ otherwise.

Example 2.3 (Down-and-out and other Barrier options)**.** *Given a strike K, maturity T and a* barrier $B > 0$, *the* down-and-out call option *has the payoff*

$$c_{DOC} := (S_T^i - K)^+ \mathbb{1}_{\mathbb{R}_+}(\min_{t\in[0,T]} S_t^i - B).$$

The payoff of an up-and-in call option *with the same strike and maturity is*

$$c_{UIC} := (S_T^i - K)^+ \mathbb{1}_{\mathbb{R}_+}(\max_{t\in[0,T]} S_t^i - B).$$

Barrier put options have similar payoffs. For example, down-and-in put options *have payoffs of the form*

$$c_{DIP} : (K - S_T^i)^+ \mathbb{1}_{\mathbb{R}_+}(B - \min_{t\in[0,T]} S_t^i).$$

Options that depend on multiple underlying assets are called *rainbow options*.

Example 2.4 (Basket options)**.** *Given a set of assets indexed by $i = 1, \dots, I$ and positive coefficients a_i, $i = 1, \dots I$, the payoff of the corresponding* basket call option *is*

$$c_{BC} := \left(\sum_{i=1}^I a_i S_T^i - K\right)^+.$$

Similarly, the basket put option *has the payoff*

$$c_{BP} := \left(K - \sum_{i=1}^I a_i S_T^i\right)^+.$$

Example 2.5 (Spread options)**.** *Given two assets* S^1 *and* S^2*, the payoff of the corresponding* spread call option *is*

$$c_{SC} := \left(S_T^1 - S_T^2 - K\right)^+ .$$

Similarly, the spread put option *has the payoff*

$$c_{SP} := \left(K - S_T^1 + S_T^2\right)^+ .$$

Example 2.6 (Calls and puts on max and min)**.** *Given to assets* S^1 *and* S^2*, the payoff of the corresponding* call-on-max option *is*

$$c_{MaxC} := \left(\max\{S_T^1, S_T^2\} - K\right)^+ .$$

Similarly, the put-on-min option *has the payoff*

$$c_{MinP} := \left(K - \min\{S_T^1, S_T^2\}\right)^+ .$$

Many options depend on quantities that are not tradable on markets.

Example 2.7 (Options on non-tradables)**.** *Let* ξ_T *be the temperature (somewhere of interest) at time* T*, and consider options with the payoffs*

$$(\xi_T - K)^+ \quad \textit{and} \quad (K - \xi_T)^+$$

with a given strike K*.*

Example 2.8 (American options*)**.** *The holder of an American option may choose to exercise the option at any time before the terminal time* T*. For example, for an American call on* S^i *with strike* K*, the payoff, if the holder chooses to exercises the option at time* t*, is*

$$(S_t - K)^+.$$

In contrast to all the above options, the holder of an American faces an optimization problem when to exercise the option.

2.2 Exercises

In all the exercises, examples in Matlab online help pages help you to write the actual code.

Exercise 2.2.1. *Write Matlab functions (as .m-files) of the payoff functions in Examples 2.1–2.6. Write them as functions of the underlying asset prices and strikes.*

Exercise 2.2.2. *Using the plot-function, plot the European call option, for a fixed strike* K*, as a function of the underlying asset price* S_T*. Plot the European call option as a function of the underlying asset price* S_T *for two different strikes in the same figure.*

Exercise 2.2.3. *Using the mesh-function (or surf-function), draw a 3D-graph of the spread call option as a function of the underlying asset prices* S_T^1 *and* S_T^2*.*

2.3 Basic properties of Brownian motion

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t=0}^T, P)$ be a filtered probability space. We consider continuous time stochastic processes only on the "time interval" $[0, T]$. A family $S := (S_t)_{t \in [0,T]}$ of $\mathbb{R}^d$–valued random variables S_t is called an $\mathbb{R}^d$–valued *continuous time stochastic process*. The process is called *adapted* if S_t is $\mathcal{F}_t$-measurable for each $t \in [0, T]$.

Given $\omega \in \Omega$, the function $t \mapsto S_t(\omega)$ is called as a *path*, or a *trajectory* or a *realization*, of the process S. Instead of considering a stochastic process as an indexed family of $\mathbb{R}^d$-valued random variables, one may thus think of a stochastic process as a family of random paths, trajectories, etc. In some cases (less in this course), it is helpful to think of a stochastic process S as a function $(\omega, t) \mapsto S_t(\omega)$ from the product space $\Omega \times [0, T]$ to $\mathbb{R}^d$. If the paths of a continuous time process are P-almost surely continuous, then the process is called *a continuous stochastic process*.

For a random variable $\eta \in (\Omega, \mathcal{F}, P)$, we denote $\eta \sim N(\mu, \sigma^2)$ when η is a normally distributed random variable with mean μ and standard deviation σ.

Remark 2.9. *We often use the property that for $\eta \sim N(0, \sigma^2)$ and positive integer m, there is a constant L such that $E\eta^{2m} = L\sigma^{2m}$,*

Definition 2.10. *An adapted continuous stochastic process W is a Brownian motion, if it has independent increments in the sense that, for all $0 \leq t_0 < t_1 < \cdots < t_n$ the random variables $\{W_{t_i} - W_{t_{i-1}} \mid i = 1, \dots n\}$ are independent, and $W_t - W_s \sim N(0, t - s)$ for all $0 \leq s < t \leq T$,*

From now on we assume, unless stated otherwise, that given a Brownian motion W, it starts at zero, that is, $W_0 = 0$.

Exercise 2.3.1. *Show that a Brownian motion W is a martingale, that is, for all $s < t \leq T$, $s > 0$, we have $E|W_t| < \infty$ and*

$$E[W_t \mid \mathcal{F}_s] = W_s.$$

Here we assume that the increments of W are independent of the filtration in the sense that, for all $s < t$, the random variable $W_t - W_s$ is indenpendent of $\mathcal{F}_s$. This is the case, .e.g., when the filtration is generated by W.

In the definition of Brownian motion, it possible to omit the assumption that the paths are continuous. This follows from the famous Kolmogorov's continuity criterion. Recall that a continuous function $f : [0, T] \to \mathbb{R}$ is α-Hölder continuous if there is $L \in \mathbb{R}$ such that

$$|f_t - f_s| \leq L|t - s|^\alpha \quad \forall\ 0 \leq s \leq t \leq T.$$

Theorem 2.11 (Kolmogorov's continuity criterion)**.** *Let S be a stochastic processes with*

$$E\,|S_t - S_s|^a \leq L\,|t - s|^{1+b} \quad \forall\ s < t \tag{2.1}$$

for some constants $a \geq 1$, $b, L > 0$. Then there exists a continuous stochastic process $\tilde{S}$ that is a modification of S in the sense that $P(\tilde{S}_t = S_t) = 1$ for all t. Moreover, $\tilde{S}$ is α-Hölder continuous almost surely for any $\alpha \leq b/a$.

Exercise 2.3.2. *Using Remark 2.9, show that, for any $\epsilon > 0$, Brownian motion has $(1/2 - \epsilon)$-Hölder continuous paths almost surely.*

From the computational perspective, Brownian motion has the important property that it can be approximated by piece-wise constant "discrete-time random walks" that have independent increments. Such random random walks are easy to simulate which is the basis of Monte Carlo methods that is the main topic of the course.

Recall that a sequence of random variables (η^ν) converges in distribution to the random variable η if

$$P(\eta^\nu \leq x) \to P(\eta \leq x)$$

for all $x \in \mathbb{R}$ such that $x \mapsto P(\eta \leq x)$ is continuous (i.e., for all x such that the cumulative distribution function of η is continuous at x). A sequence of vectors of random variables $(\eta_1^\nu, \dots \eta_k^\nu)$ converges in distribution to $(\eta_1, \dots, \eta_k)$ if

$$P((\eta_1^\nu, \dots, \eta_k^\nu) \leq x) \to P((\eta_1, \dots, \eta_k) \leq x)$$

for all $x \in \mathbb{R}^k$ such that $x \mapsto P((\eta_1, \dots, \eta_k) \leq x)$ is continuous.

Theorem 2.12 (The central limit theorem)**.** *Let*

$$\eta^{(n)} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k,$$

for an i.i.d. $(\xi_k)_{k=1}^\infty$ sequence of random variables with $E\xi_k = 0$ and $E(\xi_k)^2 = 1$. We have

$$\eta^{(n)} \xrightarrow{d} \eta \tag{2.2}$$

for a random variable $\eta \sim N(0, 1)$.

For continuous time stochastic processes $S^{(n)}$, $n = 1, 2, \dots$ and S, $S^{(n)}$ converges in finite dimensional distributions to S, denoted by

$$S^{(n)} \xrightarrow{fd} S,$$

if, for all integers k and all $0 \leq t_0 < \cdots < t_k \leq T$,

$$(S_{t_0}^{(n)}, \dots, S_{t_k}^{(n)}) \xrightarrow{d} (S_{t_0}, \dots, S_{t_k}).$$

Theorem 2.13. *Let*

$$Y_t^{(n)} := \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor nt \rfloor} \xi_k$$

for an i.i.d. $(\xi_k)_{k=1}^\infty$ *sequence of random variables with* $E\xi_k = 0$ *and* $E(\xi_k)^2 = 1$. *Then*

$$Y_t^{(n)} \xrightarrow{fd} W$$

for a Brownian motion W.

Proof. Using the central limit theorem and $\frac{\lfloor nt \rfloor}{n} \to t$ when $n \to \infty$, we get

$$Y_t^{(n)} = \frac{\sqrt{\lfloor nt \rfloor}}{\sqrt{n}} \frac{1}{\sqrt{\lfloor nt \rfloor}} \sum_{k=1}^{\lfloor nt \rfloor} \xi_k \xrightarrow{d} \eta \sim N(0,t),$$

as $n \to \infty$. Let now $t < u$. The random variables $Y_u^{(n)} - Y_t^{(n)}$ are independent from the variables $Y_t^{(n)}$, since

$$Y_u^{(n)} - Y_t^{(n)} = \frac{1}{\sqrt{n}} \sum_{k=\lfloor nt \rfloor+1}^{\lfloor nu \rfloor} \xi_k^{(n)}$$

and the random variables $\xi_k^{(n)}$ are independent. Repeating the previous arguments we get

$$Y_u^{(n)} - Y_t^{(n)} \xrightarrow{d} \eta_{u-t} \sim N(0, u-t).$$

We observe that the variables $\Delta Y_{t_i}^{(n)} := Y_{t_i}^{(n)} - Y_{t_{i-1}}^{(n)}$ are mutually independent for all $0 \le t_0 < t_1 < \cdots < t_N \le T$. Thus the process $Y^{(n)}$ has independent increments, and so

$$P\left(Y_{t_i}^{(n)} - Y_{t_{i-1}}^{(n)} \le x_i, i = 1, \dots, N\right) = \prod_{i=1}^N P\left(Y_{t_i}^{(n)} - Y_{t_{i-1}}^{(n)} \le x_i\right)$$
$$\longrightarrow \prod_{i=1}^N \Phi_{0,t_i - t_{i-1}}(x_i) = P(W_{t_i} - W_{t_{i-1}} \le x_i, i = 1, \dots, N).$$

The proof is finished by the next exercise. □

Exercise 2.3.3. *Recall the continuous mapping theorem: If* $(\eta_0^\nu, \dots, \eta_k^\nu) \xrightarrow{d} (\eta_0, \dots, \eta_k)$, *then* $f(\eta_0^\nu, \dots, \eta_k^\nu) \xrightarrow{d} f(\eta_0, \dots, \eta_k)$ *for any continuous function* $f : \mathbb{R}^k \to \mathbb{R}^n$.

Use the continuous mapping theorem to finish the proof of Theorem 2.13.

2.4 Exercises

In all the exercises, examples in Matlab online help pages help you to write the actual code. We say that the process defined by

$$B_t := \mu t + \sigma W_t$$

is a Brownian motion with drift μ and volatility σ. Here W is a (standard) Brownian motion.

Exercise 2.4.1. *Write a Matlab function (as an .m-file) that creates a sample path of a Brownian motion with terminal time T, $n+1$ equi-distant discretization points, drift μ and volatility σ. Write it as a function of these parameters and a sample of independent standard normally distributed random variable so that the function maps the sample to a (discretized) sample path of a Brownian motion.*

Exercise 2.4.2. *Plot sample paths of the Brownian motion with different drifts and volatilities in the same figure.*

Download the .mat files from the course page. They contain "classes" consisting of sample paths of a Brownian motion with a given terminal time T.

Exercise 2.4.3. *For paths in "bmpaths.mat", estimate the volatility of each path. Plot the paths in the same figure and label the paths with their volatilities.*

2.5 Quadratic variation

Let

$$D := \bigcup_n D_n,$$

where D_n is the *n-th dyadic partition* of $[0,T]$,

$$D_n := \{i/2^n \in [0,T] \mid i = 0,1,2\dots,\}.$$

Enumerating $D_n = \{\{t_0^n, t_1^n, \dots\} \mid t_i^n \le t_{i+1}^n\}$, we define, for each n, the "discrete quadratic variation" of a stochastic process S by

$$QV_t^n(S) := \sum_{i\ge 1} |S_{t_i^n\wedge t} - S_{t_{i-1}^n\wedge t}|^2,$$

where $s \wedge t := \min\{s,t\}$.

Theorem 2.14. *Let W be a Brownian motion. Then*

$$P(\lim_n QV_t^n(W) = t \quad \forall t) = 1.$$

Proof. Fix $t \in D$. The almost sure convergence $QV_t^n(W) \to t$ is equivalent to the almost sure convergence $\sum_{t_n^i\le t} Z_i \to 0$ for

$$Z_i := (W_{t_{i+1}^n\wedge t} - W_{t_i^n\wedge t})^2 - 2^{-n}.$$

Note first that $t_{i+1}^n - t_i^n = 2^{-n}$ so that $W_{t_{i+1}^n\wedge t} - W_{t_i^n\wedge t}$ are mutually independent and $N(0,2^{-n})$-distributed. We have $E[Z_iZ_j] = 0$ for $i \ne j$ and $EZ_i^2 = L2^{-2n}$

for some constant L. Using the monotone convergence theorem, we get

$$\begin{aligned}E\sum_{n\geq 1}(\sum_{t_i^n\leq t} Z_i)^2 &= \lim_{N\to\infty}\sum_{n\leq N} E(\sum_{t_i^n\leq t} Z_i)^2 \\ &= \lim_{N\to\infty}\sum_{n\leq N}(\sum_{t_i^n\leq t} L2^{-2n}) = L\sum_{n\geq 1} t2^{-n} < +\infty.\end{aligned}$$

Therefore $\sum_{n\geq 1}(\sum_{t_i^n\leq t} Z_i)^2$ is almost surely finite, and thus $(\sum_{t_i^n\leq t} Z_i)^2$ converge to zero (if an infinite sum of real numbers convergences, then the summands have to converge to zero). But then also $\sum_{t_i^n\leq t} Z_i$ converges to zero, so we have shown that $QV_t^n(W)\to t$ almost surely.

Since D is countable, we can find a P-null-set N such that $QV_s^n(W)\to s$ for every $s\in D$ and $\omega\notin N$. Since, for each n, $t\mapsto QV_t^n(W)$ is increasing, we get, for $s^\nu\in D$ increasing to t, $t^\nu\in D$ decreasing to t, and for every $\omega\notin N$, that

$$t=\lim_\nu s^\nu = \lim_\nu\lim_n QV_{s^\nu}^n(W)\leq \lim QV_t^n(W)\leq \lim_\nu\lim_n QV_{t^\nu}^n(W)=\lim_\nu t^\nu = t$$

□

Lemma 2.15. *Assume that z is an adapted continuous stochastic process with $\sup_t Ez_t^2<\infty$. Then, for every t,*

$$\lim\sum_{t_i^{(n)}\leq t}(z_{t_i^{(n)}}(W_{t_{i+1}^{(n)}}-W_{t_{i+1}^{(n)}})^2)=\int_0^t z_s ds,$$

where the convergence is in L^2.

Proof. We denote $\eta_i := z_{t_i^{(n)}}$, $L=\sup_t Ez_t^2<\infty$, $\Delta t_{i+1}^{(n)} := t_{i+1}^{(n)}-t_i^{(n)}$ and $\Delta W_{i+1}^{(n)} := W_{t_{i+1}^{(n)}}^{(n)}-W_{t_i^{(n)}}^{(n)}$ so that

$$\sum_{t_i^{(n)}\leq t}\left(z_{t_i^{(n)}}(W_{t_{i+1}^{(n)}}-W_{t_i^{(n)}})^2-z_{t_i^{(n)}}(t_{i+1}^{(n)}-t_i^{(n)})\right)=\sum_{t_i^{(n)}\leq t}\eta_i((\Delta W_{i+1}^{(n)})^2-\Delta t_{i+1}^{(n)}).$$

Recalling that the increments of Brownian motion are independent of the past and z is adapted, we get from $E[(\Delta W_{i+1}^{(n)})^2-\Delta t_{i+1}^{(n)})]=0$ and independence that

$$\begin{aligned}&E\left[\eta_i(\Delta W_{i+1}^{(n)})^2-\Delta t_{i+1}^{(n)})\eta_j(\Delta W_{j+1}^{(n)})^2-\Delta t_{j+1}^{(n)})\right]\\ =&E\left[(\Delta W_{i+1}^{(n)})^2-\Delta t_{i+1}^{(n)})\right]\left[E\eta_i\eta_j(\Delta W_{j+1}^{(n)})^2-\Delta t_{j+1}^{(n)})\right]\\ =&0\end{aligned}$$

for $i > j$. Combining with Remark 2.9, we get for some constants L (differing from line to line),

$$\begin{aligned} E|\sum_{t_i^{(n)} \le t} \eta_i((\Delta W_{i+1}^{(n)})^2 - \Delta t_{i+1}^{(n)})|^2 &= \sum_{t_i^{(n)} \le t} E|\eta_i((\Delta W_{i+1}^{(n)})^2 - \Delta t_{i+1}^{(n)})|^2 \\ &\le L \sum_{t_i^{(n)} \le t} E((\Delta W_{i+1}^{(n)})^2 - \Delta t_{i+1}^{(n)})^2 \\ &\le L \sum_{t_i^{(n)} \le t} \left(\Delta t_{i+1}^{(n)}\right)^2 \\ &= Lt2^n(2^{-2n}) \\ &\to 0. \end{aligned}$$

Since $\sum_{t_i^{(n)} \le t} z_{t_i^{(n)}} \Delta t_{i+1}^n$ converges to $\int_0^t z_s ds$ in L^2, the claim follows from the triangle inequality. $\square$

Remark 2.16. *Choosing $z = 1$ in Lemma 2.15, we get $QV^n(W)_t \to t$ in L^2.*

2.6 Stochastic integrals

Theorem 2.14 implies that the paths of the Brownian motion are not of bounded variation, and thus not differentiable. Indeed,

$$QV_t^n(W) \le \max_{i \ge 1} |W_{t_{i+1}^{(n)}} - W_{t(n)_i}| \sum_{i \ge 1} |W_{t(n)_{i+1}} - W_{t(n)_i}|,$$

where, almost surely, $QV_t^n(W)$ converge to t and $\max_{i \ge 1} |W_{t(n)_{i+1}} - W_{t(n)_i}|$ converges to zero (by continuity of BM), so $\sum_{i \ge 1} |W_{t(n)_{i+1}} - W_{t(n)_i}|$ has to converge to $+\infty$. This means that it is not possible integrate functions with respect to the paths of Brownian motion in the usual sense of the Lebesque-Stieltjes integration theory.

However, it is possible to define integrals with respect to the Brownian motion in the sense of stochastic integrals. An adapted stochastic process z is *simple* if

$$z_t = \sum_{i=0}^{\infty} \eta_i 1_{(t_i, t_{i+1}]}(t)$$

for some $0 \le t_1 \le t_2 \le \dots$ and $\mathcal{F}_{t_i}$-measurable η_i with $\sup_i \operatorname{ess\,sup} |\eta_i| < \infty$. For a simple z, we set

$$\int_0^t z_t dW_t := \sum_{i=0}^{T} \left(\eta_i (W_{t \wedge t_{i+1}} - W_{t \wedge t_i})\right).$$

We extend the definition from simple processes to larger spaces of integrands

$$\mathcal{H}^2 := \{z \mid z \text{ measurable adapted stochastic process}, E\int_0^T |z_t|^2 dt < \infty\},$$

which we equip with the norm $\|z\|_{\mathcal{H}^2} := (E\int_0^T |z_t|^2 dt)^{1/2}$. For $z \in \mathcal{H}^2$, we define the stochastic integral as the unique limit

$$\int_0^t z_s dW_s := \lim_n \int_0^t z_s^{(n)} dW_s$$

in L^2, where $(z^{(n)})$ is any sequence of simple processes converging to z in $\mathcal{H}^2$.

Example 2.17. *Let $z = 2W$. For $\{t_0^n, t_i^n, \dots\} = D^n$, it is possible to show that the processes $z_t^{(n)} = \sum_{i\geq 0} 2W_{t_i^n} 1_{(t_i^n, t_{i+1}^n]}(t)$ converge to W in $\mathcal{H}^2$. We have*

$$\begin{aligned}\int_0^t z_s^{(n)} dW_s &:= \sum_{i\geq 0}\Big(2W_{t_i^n}(W_{t\wedge t_{i+1}^n} - W_{t\wedge t_i^n})\Big)\\ &= \sum_{i\geq 1}(W^2_{t\wedge t_{i+1}^n} - W^2_{t\wedge t_i^n}) - \sum_{i\geq 1}(W_{t\wedge t_{i+1}^n} - W_{t\wedge t_i^n})^2\\ &= W_t^2 - QV_t^{(n)}(W)\\ &\to W_t^2 - t,\end{aligned}$$

where the convergence is in L^2, by Remark 2.16. Thus

$$\int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{1}{2}t.$$

Theorem 2.18. *Let $z \in \mathcal{H}^2$ and S be the stochastic process defined by*

$$S_t = \int_0^t z_s dW_s.$$

1. *The process S is a continuous martingale that belongs to $\mathcal{H}^2$,*
2. *We have the* Itô isometry $ES_T^2 = \|z\|^2_{\mathcal{H}^2}$,
3. *If z is deterministic (and $\int_0^T |z_s|^2 ds < \infty$), then S has independent increments and $(S_t - S_s) \sim N(0, \int_s^t |z_u|^2 du)$.*

Next we extend the definition of the stochastic integral to integrands in the space $\mathcal{H}^2_{\text{loc}}$, where

$$\mathcal{H}^p_{\text{loc}} := \{z \mid z \text{ measurable adapted stochastic process}, \int_0^T |z_t|^p dt < \infty \ P\text{-a.s.}\}.$$

For $z \in \mathcal{H}^2_{\text{loc}}$, we define the stochastic integral as the unique limit

$$\int_0^t z_s dW_s := \lim_n \int_0^t z_s^{(n)} dW_s$$

where $z_t^{(n)} = z_t \mathbb{1}_{t \wedge \tau^{(n)}}$ and $\tau^{(n)} = \inf_t \{\int_0^t |z_t|^2 dt \geq n\}$ (here $\tau^{(n)}$ is a "localizing sequence of z). The stochastic process defined via $\int_0^t z_t dW_t$ is a continuous process, but not necessarily a martingale (it is only a "local martingale").

2.7 Exercises

Exercise 2.7.1. *Show that a Brownian motion W is a martingale with respect to its natural filtration $\mathcal{F}_t = \sigma(W_s \mid s \leq t)$, that is, for all $s < t \leq T$, $s \geq 0$, we have*

$$E[W_t \mid \mathcal{F}_s] = W_s.$$

Exercise 2.7.2. *Using Remark 2.9, show that, for any $\epsilon > 0$, Brownian motion has $(1/2 - \epsilon)$-Hölder continuous paths almost surely.*

Exercise 2.7.3. *Recall the continuous mapping theorem: If $(\eta_0^\nu, \ldots, \eta_k^\nu) \xrightarrow{d} (\eta_0, \ldots, \eta_k)$, then $f(\eta_0^\nu, \ldots, \eta_k^\nu) \xrightarrow{d} f(\eta_0, \ldots, \eta_k)$ for any continuous function $f : \mathbb{R}^k \to \mathbb{R}^n$.*

Use the continuous mapping theorem to finish the proof of Theorem 2.13.

Exercise 2.7.4. *In the setting of Example 2.17, show that $z^{(n)} \to z$ in $\mathcal{H}^2$.*

Exercise 2.7.5. *Let S be a stochastic process defined by*

$$S_t = \exp\{\int_0^t z_s dW_s - \frac{1}{2} \int_0^t |z_s|^2 ds\},$$

where z is deterministic with $\int_0^T |z_s|^2 ds < \infty$.

- *Compute $E \exp(\eta)$ for $\eta \sim N(0, \sigma^2)$, where $\sigma \in \mathbb{R} > 0$.*
- *Show that S is a martingale without relying on the first part of Theorem 2.18.*

2.8 Itô processes and Itô's formula

An important difference to the classical integration theory is that the stochastic integral does not satisfy the usual chain rule. Recall that for continuously differentiable functions g on $\mathbb{R}$ and f on $[0, T]$, we have $\frac{d}{dt} g(f) = g'(f) f'$ and so

$$g(f_t) = g(f_0) + \int_0^t g'(f_s) df_s.$$

Example 2.17 shows that this is not the case for the stochastic integral, since we got

$$\frac{1}{2}(W_t)^2 = \int_0^t W_s dW_s + \frac{1}{2}t,$$

where we have an "Itô correction term" involving the quadratic variation of W. This observation generalizes to the famous Ito's formula that we formulate directly to Itô processes.

Definition 2.19. *A stochastic process X is called an* Itô process*, if there is $\mu \in \mathcal{H}^1_{loc}$ and $\sigma \in \mathcal{H}^2_{loc}$ such that*

$$X_t = X_0 + \int_0^t \mu_s ds + \int \sigma_s dW_s.$$

The definition of the stochastic integral extends to Itô processes. Let

$$X_t = X_0 + \int_0^t \mu_s ds + \int \sigma_s dW_s$$

be an Itô process with $\mu \in \mathcal{H}^1_{\text{loc}}$ and $\sigma \in \mathcal{H}^2_{\text{loc}}$. For any z such that $z\mu \in \mathcal{H}^1_{\text{loc}}$ and $z\sigma \in \mathcal{H}^2_{\text{loc}}$, we define

$$\int_0^t z_s dX_s := \int_0^t z_s \mu_s ds + \int_0^t z_s \sigma_s dW_s.$$

We denote by $C^{1,2}$ the continuous functions $(t,x) \mapsto f(t,x)$ on $[0,T] \times \mathbb{R}$ that are continuously differentiable once w.r.t t and twice w.r.t. x.

Theorem 2.20 (Itô's formula)**.** *Assume that $f \in C^{1,2}([0,T] \times \mathbb{R})$ and that*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

for some $\mu \in \mathcal{H}^1_{loc}$ and $\sigma \in \mathcal{H}^2_{loc}$. Then, almost surely,

$$f(t, X_t) = f(0, X_0) + \int_0^t \partial_x f(s, X_s) dX_s + \int_0^t (\partial_t f(s, X_s) + \frac{1}{2}\sigma_s^2 \partial_{xx} f(s, X_s)) ds.$$

Proof. We do not give the whole proof, but we only demonstrate how the "correction term" $\frac{1}{2}\sigma_s^2 \partial_{xx} f(s, X_s))ds$ appears to the formula in the special case when $X = W$, f is constant w.r.t. t-component, $f(0) = 0$, and $\partial_x f$ and $\partial_{xx} f$ are bounded. For general X, the argument follows similarly while boundedness of the derivatives can be handled using localizing sequences of X. Using Taylor's expansion (below $\eta_i^{(n)}$ is the appropriate random variable with

$W_{t_{i+1}^{(n)}} \leq \eta_i^{(n)} \leq W_{t_{i+1}^{(n)}}$), we get

$$
\begin{aligned}
(f(W_{t_{i+1}^{(n)}} - f(W_{t_i^{(n)}})) = \partial_x f(W_{t_i^{(n)}})(W_{t_{i+1}^{(n)}} - (W_{t_i^{(n)}}) \\
+ \frac{1}{2}\partial_{xx} f(W_{t_i^{(n)}})(W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}})^2 \\
+ \frac{1}{2}(\partial_{xx} f(\eta_i^{(n)}) - \partial_{xx} f(W_{t_i^{(n)}}))(W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}})^2.
\end{aligned}
$$

Summing over i we arrive at

$$
\begin{aligned}
f(W_t) = & \sum_n \partial_x f(W_{t_i^{(n)}})(W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}}) \\
& + \sum_n \frac{1}{2}\partial_{xx} f(W_{t_i^{(n)}})(W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}})^2 \\
& + \sum_n \frac{1}{2}(\partial_{xx} f(\eta_i^{(n)}) - \partial_{xx} f(W_{t_i^{(n)}}))(W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}})^2.
\end{aligned}
$$

As n tends to infinity, the first sum converges to $\int_0^t \partial_x f(W_s) dW_s$, the second sum converges to $\int_0^t \partial_{xx} f(W_s) ds$, by Lemma 2.15, and the last term tends to zero (for this we omit the details). $\square$

We will also need the following (local) martingale representation result.

Theorem 2.21 (The martingale representation theorem)**.** *Assume that $\eta \in L^1(\mathcal{F}_T)$ and that the filtration is generated by a Brownian W. Then there exists $z \in \mathcal{H}^2_{loc}$ such that, almost surely,*

$$
\eta = E[\eta] + \int_0^T z_t dW_t.
$$

If $\eta \in L^2(\mathcal{F}_T)$, then $z \in \mathcal{H}^2$ and z in the above representation is unique.

2.9 Multi-dimensional results

We denote the $\mathbb{R}^{d\times d}$-identity matrix by 1_d.

Definition 2.22. *An $\mathbb{R}^d$-valued stochastic process W is a d-*dimensional standard Brownian motion *if the components W^i are independent standard Brownian motions and $W_t - W_s \sim N(0, (t-s)1_d)$.*

Theorem 2.23. *Let W be a d-dimensional standard Brownian motion. Then*

$$
\sum_{t_i^n \leq t} (W_{t_{i+1}^n} - W_{t_t^n})(W_{t_{i+1}^n} - W_{t_t^n})^T \xrightarrow{n\to\infty} t1_d
$$

where the convergence is in L^2.

Definition 2.24. *A stochastic process X is called a d-dimensional* Itô process*, if there is $\mu \in \mathcal{H}^1_{loc}(\mathbb{R}^d)$ and $\sigma \in \mathcal{H}^2_{loc}(\mathbb{R}^d \times \mathbb{R}^n)$ such that*

$$X^i_t = X^i_0 + \int_0^t \mu^i_s ds + \int \sigma^i_s dW_s,$$

where σ^i_s is the i-th row of the matrix σ_s and W is an n-dimensional standard Brownian motion.

Theorem 2.25 (Multi-dimensional Itô's formula)**.** *Assume that $f \in C^{1,2}([0,T] \times \mathbb{R}^d)$ and that*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_t dW_s$$

for some a n-dimensional standard Brownian motion and $\mu \in \mathcal{H}^1_{loc}(\mathbb{R}^d)$ and $\sigma \in \mathcal{H}^2_{loc}(\mathbb{R}^{d\times n})$. Then, almost surely,

$$f(t, X_t) = f(0, X_0) + \int_0^t \partial_x f(s, X_s) dX_s + \int_0^t (\partial_t f(s, X_s) + \frac{1}{2} \operatorname{Tr}[\partial_{xx} f(s, X_s) \sigma_s \sigma_s^T]) ds.$$

Here

$$\int_0^t \partial_x f(s, X_s) dX_s := \sum_{i=1}^d \int_0^t \partial_{x^i} f(s, X_s) dX^i_s.$$

Exercise 2.9.1 (Integration by parts formula for Itô processes)**.** *Let W be one dimensional Brownian motion and let X^i, $i = 1, 2$ be one dimensional Itô processes, i.e.,*

$$X^i_t = X^i_0 + \int_0^t \mu^i_s ds + \int_0^t \sigma^i_s dW_s \quad i = 1, 2,$$

where, for $i = 1, 2$, $\mu^i \in \mathcal{H}^1_{loc}$ and $\sigma^i \in \mathcal{H}^2_{loc}$. Apply Itô's formula to $f(t, (x^1, x^2)) = x^1 x^2$ to show the integration by parts formula

$$X^1_t X^2_t = X^1_0 X^2_0 + \int_0^t X^2_s dX^1_s + \int_0^t X^1_s dX^2_s + \int_0^t \sigma^1_s \sigma^2_s ds.$$

Exercise 2.9.2. *Let W be a d-dimensional Brownian motion, $b \in \mathbb{R}$ and $\sigma \in \mathbb{R}^d$, and consider the process*

$$S_t = \exp\{bt + \sigma \cdot W_t\}.$$

For $f \in C^{1,2}([0,T] \times \mathbb{R})$, show that $f(t, S_t)$ is an Itô process. Can you find a function f such that the process given by $Y_t := f(t, S_t)$ is a local martingale?

2.10 Exercises

We say that the process defined by

$$B_t := \mu t + \sigma W_t$$

is a Brownian motion with drift $\mu \in \mathbb{R}^d$ and volatility $\sigma \in \mathbb{R}^{d\times d}$, where W is a (standard) d-dimensional Brownian motion W. The matrix $Cov = \sigma\sigma^T$ is called the *covariation matrix*.

Instead of specifying σ, it is more common to specify Cov which can be estimated from data. It is an exercise to check that volatility matrices with a common covariance matrix define Brownian motions with common finite dimensional distributions.

Example 2.26. *Assume that $\eta = (\eta_1, \dots, \eta_d)$ is a vector of independent standard normally distributed random variables and let $\Sigma \in \mathbb{R}^{d\times d}$ be given. Let $\sigma \in \mathbb{R}^{d\times d}$ be such that $\sigma\sigma^T = \Sigma$. Then*

$$\mathrm{Cov}(\sigma\eta) = E[(\sigma\eta)(\sigma\eta)^T] = \sigma E[\eta\eta T]\sigma^T = \sigma\sigma^T = \Sigma.$$

Thus the covariance depends on σ only through $\sigma\sigma^T$. We can always use the Cholesky decomposition $\Sigma = LL^T$, where L is a triangular matrix.

For B_t this implies the following. When only the covariance is specified, one can always use L from the Cholesky decomposition as σ. Given any other $\hat{\sigma}$ that gives the same covariance, the processes B and

$$\hat{B}_t := \mu t + \hat{\sigma} W_t$$

have the same finite dimensional disributions.

Exercise 2.10.1. *Write a Matlab function (as an .m-file) that creates a sample path of a d-dimensional Brownian motion with terminal time T, $n+1$ equi-distant discretization points, drift vector μ and covariation matrix Cov. Write it as a function of these parameters and an i.i.d. sample of d-dimensional standard normals. Hint: Use Cholesky decomposition of Cov.*

Exercise 2.10.2. *Plot a sample path of a 2-dimensional Brownian motion with terminal time T, $n+1$ equi-distant discretization points with a drift vector μ and covariation matrix Cov. Plot it as a 3D graph, a 2D parametric curve (time being the parameter), and as each component as a different curve in the same figure. Which plot is the most informative?*

Exercise 2.10.3. *Plot sample paths of a 3-dimensional Brownian motion with terminal time T, $n+1$ equi-distant discretization points with a drift vector μ and with different covariation matrices Cov. Plot it so that each component is a different curve in the same figure. Vary Cov so that the role of covariance matrix becomes clear in the figures.*

Exercise 2.10.4. *For the path in "bmpath3D.mat" of a 3D Brownian motion, estimate the covariation matrix. Plot the components of the path in the same figure and label the paths with the rows of the covariance matrix (with the precision of two digits).*

3 The Black-Scholes model

Let $W = (W_t^i, \ldots, W_t^d)_{t\in[0,T]}$ be a d-dimensional standard Brownian motion. The financial market consists of $d+1$ assets. The asset S^0 is a "non-risky" asset defined by

$$S_t^0 = e^{rt},$$

where r models the instantaneous interest rate. The risky assets are modelled by

$$S_t^i = S_0^i \exp\left(\left(\mu^i - \frac{1}{2}\sum_{j=1}^d |\sigma^{ij}|^2\right)t + \sum_{j=1}^d \sigma^{ij} W_t^j\right),$$

where S_0^i are the initial prices, and μ^i and σ^{ij} are constants, describing "drifts" and correlations between the assets, respectively. We assume that the matrix formed by σ^{ij} is invertible.

Example 3.1. *Consider a model with only one risky asset with $S_0^1 = 1$. Omitting indices from S^1, W^i and from the parameters μ^1 and σ^{11}, the model of the risky asset becomes*

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

Defining

$$f(t,x) = \exp\left(\left(\mu - \frac{1}{2}|\sigma|^2\right)t + \sigma x\right),$$

we have $S_t = f(t, W_t)$, so Itô's formula gives

$$\begin{aligned}
S_t &= f(t, W_t) \\
&= f(0,0) + \int_0^t \partial_x f(s, W_s)dW_s + \int_0^t (\partial_t f(s,W_s) + \frac{1}{2}\partial_{xx} f(s, W_s))ds \\
&= 1 + \int_0^t \sigma S_s dW_s + \int_0^t (\left(\mu - \frac{1}{2}|\sigma|^2\right) S_s + \frac{1}{2}\sigma^2 S_s)ds \\
&= 1 + \int_0^t \sigma S_s dW_s + \int_0^t \mu S_s ds.
\end{aligned}$$

Therefore, S solves the "stochastic differential equation" ("SDE")

$$dS_t = S_t(\mu dt + \sigma dW_t) \quad S_0 = 1.$$

Applying Itô's formula to the d-dimensional model, just as in Example 3.1, we see that the risky assets solve the SDE

$$dS_t^i = S_t^i(\mu^i dt + \sum_{j=1}^d \sigma^{ij} dW_t^j), \quad S_0^i = 1, \quad \forall\ i = 1, \ldots, d.$$

Using the notations $\mu = (\mu^1, \ldots, \mu^d))$, $\sigma \in \mathbb{R}^{d\times d}$ with entries σ^{ij}, and $\mathrm{diag}[S_t]$ for the diagonal $\mathbb{R}^{d\times d}$-matrix with entries S_t^i, this can be written as

$$dS_t = \mathrm{diag}[S_t](\mu dt + \sigma dW_t), \quad S_0 = \mathbb{1}. \tag{3.1}$$

A portfolio process $\theta = (\theta_t)_{t\in[0,T]}$ is an adapted $\mathbb{R}^d$-valued stochastic process. The number θ_t^i describes the amount in Euros invested in the i-th risky asset at time t, so the ratio $z_t^i := \frac{\theta_t^i}{S_t^i}$ is the amount of i-th asset held in the portfolio at time t.

Let $X^\theta = (X_t^\theta)_{t\in[0,T]}$ denote the $\mathbb{R}$-valued stochastic process describing the wealth accumulated by the portfolio process θ. Then the amount invested in the non-risky asset at time t is $\theta_t^0 := X_t^\theta - \sum_{i=1}^d \theta_t^i = X_t^\theta - \mathbb{1}\cdot\theta_t$.

When each z_t^i is a piecewise constant (i.e., simple) process $z_t^i = \sum_{k=0}^\infty z_{t_k}^i 1_{(t_k,t_{k+1}]}(t)$, $0 \le t_0 \le t_1 \le \ldots$, $z_{t_k}^i \in \mathcal{F}_{t_k}$, the "self-financing condition" means that

$$X_{t_{K+1}}^\theta = \sum_{k\le K}\left(\sum_{i=1}^d z_{t_k}^i (S_{t_{k+1}}^i - S_{t_k}^i) + z_{t_k}^0 (S_{t_{k+1}}^0 - S_{t_k}^0)\right),$$

i.e., the wealth X^θ is generated solely by the portfolio process θ.

For a general θ, the *self-financing condition* is defined by

$$dX_t^\theta = \sum_{i=1}^d \frac{\theta_t^i}{S_t^i} dS_t^i + \frac{X_t - \mathbb{1}\cdot\theta_t}{S_t^0} dS_t^0 \tag{3.2}$$

as soon as the stochastic integral is well-defined. Recalling the formula (3.1) and that $S_t^0 = e^{rt}$, the self-financing condition can be written as

$$dX_t^\theta = \theta_t(\mu dt + \sigma dW_t) + r(X_t^\theta - \mathbb{1}\cdot\theta_t)dt.$$

We assume throughout that wealth processes X^θ are self-financing.

3.1 Discounted processes and the change of measure

Students familiar with "Finanzmathematik I" may recall that "discounted price processes" play an important role in the pricing theory. To this end, we define the *risk premium*

$$\lambda := \sigma^{-1}(\mu - r\mathbb{1})$$

and the *discounted price process*

$$\tilde{S}_t := \frac{S_t}{S_t^0}.$$

Exercise 3.1.1. *Prove that $\tilde{S}_t$ satisfies the SDE*

$$d\tilde{S}_t = \mathrm{diag}[\tilde{S}_t]\sigma(\lambda dt + dW_t).$$

Let $\tilde{X}^\theta$ be the discounted wealth process

$$\tilde{X}^\theta_t := X^\theta_t / S^0_t = e^{-rt} X^\theta_t.$$

Likewise, we denote by $\tilde{\theta}$ the process

$$\tilde{\theta}^i_t := \theta^i_t / S^0_t.$$

Example 3.2. *Consider the case $d = 1$. Applying Itô's formula to $f(t,x) = e^{-rt}x$, and recalling*

$$\begin{aligned} dX^\theta_t &= \frac{\theta_t}{S_t} dS_t + \frac{X_t - \theta_t}{S^0_t} dS^0_t, \\ dS_t &= S_t(\mu dt + \sigma dW_t), \end{aligned}$$

and the risk premium $\lambda = \frac{\mu - r}{\sigma}$ and $S^0_t = e^{rt}$, we get that

$$\begin{aligned} d\tilde{X}^\theta_t &= -r\frac{X^\theta_t}{S^0_t} dt + \frac{1}{S^0_t}(dX^\theta_t) \\ &= -r\frac{X^\theta_t}{S^0_t} dt + \frac{1}{S^0_t}\left(\frac{\theta_t}{S_t}(S_t(\mu dt + \sigma dW_t) + r\frac{X^\theta_t - \theta_t}{S^0_t} S^0_t dt\right) \\ &= \tilde{\theta}_t((\mu - r)dt + \sigma dW_t) \\ &= \tilde{\theta}_t \sigma(\lambda dt + dW_t). \end{aligned}$$

Recalling that $d\tilde{S}_t = \tilde{S}_t \sigma(\lambda dt + dW_t)$, this can be written as

$$d\tilde{X}^\theta_t = \frac{\tilde{\theta}_t}{\tilde{S}_t} d\tilde{S}_t.$$

The above example generalizes to the multidimensional setting and we get

$$\begin{aligned} d\tilde{X}_t &= \tilde{\theta}_t \operatorname{diag}[\tilde{S}_t]^{-1} d\tilde{S}_t \\ &= \tilde{\theta}\sigma(\lambda dt + dW_t). \end{aligned}$$

This means that the discounted wealth process is a stochastic integral of $\tilde{\theta}$ with respect to the Itô process $dX_t = \sigma(\lambda dt + dW_t)$.

Next our aim is to show that the discounted price process is a martingale under some another probability measure Q. For $Q \ll P$, the stochastic process q defined by

$$q_t := E\left[\frac{dQ}{dP} \;\middle|\; \mathcal{F}_t\right]$$

is called the *density process* of Q (with respect to P).

Lemma 3.3. *Let $Q \ll P$ and q be the density process of Q. For any $\eta \in L^1(Q)$ and $t \in [0,T]$, we have*

$$E^Q[\eta \mid \mathcal{F}_t] = \frac{1}{q_t} E\left[q_T \eta \mid \mathcal{F}_t\right] \quad Q\text{-a.s.}$$

Theorem 3.4. *Let W be a d-dimensional Brownian motion and h be a deterministic $\mathbb{R}^d$-valued measurable function on $[0,T]$ satisfying $\int_0^T |h_t|^2 dt < \infty$. Let Q be an equivalent probability measure to P with the Radon-Nikodym density*

$$dQ/dP = \exp\{\int_0^T h_t dW_t - \frac{1}{2}\int_0^T |h_t|^2 dt\}.$$

Then the stochastic process B given by

$$B_t := W_t - \int_0^t h_s ds$$

is a Brownian motion under Q.

Proof. Evidently, B is a continuous stochastic process. Thus we need to show that B has independent increments and $(B_t - B_s) \sim N(0, t-s)$ under Q for all $0 \le s < t \le T$. By Exercise 2.7.5,

$$\exp\left\{\int_0^t h_s dW_s - \frac{1}{2}\int_0^t |h_s|^2 ds\right\}$$

defines a martingale, so $q_t := E[dQ/dP \mid \mathcal{F}_t]$ satisfies

$$q_t = \exp\left\{\int_0^t h_s dW_s - \frac{1}{2}\int_0^t |h_s|^2 ds\right\}.$$

Given $\lambda \in \mathbb{R}$, we have, using Lemma 3.3,

$$\begin{aligned} E^Q[e^{\lambda(B_t - B_s)} \mid \mathcal{F}_s] &= e^{-\lambda \int_s^t h_u du} E^Q[e^{\lambda(W_t - W_s)} \mid \mathcal{F}_s] \\ &= \frac{e^{-\lambda \int_s^t h_u du}}{q_s} E[q_T e^{\lambda(W_t - W_s)} \mid \mathcal{F}_s] \\ &= e^{-\lambda \int_s^t h_u du - \frac{1}{2}\int_s^t |h_s|^2 ds} E[e^{\int_s^t (h_s + \lambda) dW_s} \mid \mathcal{F}_s]. \end{aligned}$$

By Theorem 2.18, $Y_t := \int_0^t (h_s + \lambda) dW_s$ has independent increments and $Y_t - Y_s \sim N(0, \int_s^t |h_u + \lambda|^2 du)$. Thus,

$$\begin{aligned} E[e^{\int_s^t (h_s+\lambda) dW_s} \mid \mathcal{F}_s] &= E[e^{\int_s^t (h_s+\lambda) dW_s}] \\ &= e^{\frac{1}{2}\int_s^t |h_u + \lambda|^2 du}, \end{aligned}$$

where the last line follows from Exercise 2.7.5. Combining the equalities,

$$\begin{aligned} E^Q[e^{\lambda(B_t - B_s)} \mid \mathcal{F}_s] &= e^{-\lambda \int_s^t h_u du - \frac{1}{2}\int_s^t |h_s|^2 ds} e^{\frac{1}{2}\int_s^t |h_u + \lambda|^2 du} \\ &= e^{\frac{1}{2}\lambda^2 (t-s)}. \end{aligned}$$

Thus $(B_t - B_s)$ is independent of $\mathcal{F}_s$ and the Laplace transform of $B_t - B_s$ under Q at λ equals $e^{\frac{1}{2}\lambda^2(t-s)}$. This means that $B_t - B_s \sim N(0, t-s)$ under Q. $\square$

Theorem 3.4 implies that

$$B_t := W_t + \lambda t \tag{3.3}$$

is a Brownian motion under the measure Q with

$$dQ/dP = e^{-\lambda W_T - \frac{1}{2}\lambda^2 T}. \tag{3.4}$$

We can write the the price process as

$$S_t^i = S_0^i \exp\left(\left(r - \frac{1}{2}\sum_{j=1}^{d} |\sigma^{ij}|^2\right) t + \sum_{j=1}^{d} \sigma^{ij} B_t^j\right) \tag{3.5}$$

so that, just like in Example 3.1, S solves the SDE (w.r.t B)

$$d\hat{S}_t = \operatorname{diag}[\hat{S}_t](rdt + \sigma dB_t), \quad \hat{S}_0 = S_0. \tag{3.6}$$

The discounted price process $\tilde{S}$ satisfies

$$d\tilde{S}_t = \operatorname{diag}[\tilde{S}_t]\sigma dB_t$$

while the discounted wealth process can be written as

$$\tilde{X}_t^\theta = \tilde{X}_0 + \int_0^t \tilde{\theta}_t \sigma dB_u. \tag{3.7}$$

Remark 3.5. *It is possible to show that Q is the only probability measure equivalent to P such that the discounted price process is a martingale under Q. In financial terms, this is equivalent to saying that the Black scholes market model is* complete.

Definition 3.6. *The portfolio process is called* admissible *if $\theta\sigma \in \mathcal{H}^2_{loc}$ and there exists a Q-martingale M such that such that $X_t^\theta \geq M_t$ for all t.*

Here we require the "credit limit" given in terms of the martingale M so that we do not allow "doubling strategies". We omit the detailed discussion of this pathology of continuous time market models.

4 The superhedging pricing formula and hedging

We define the superhedging price of a claim c as

$$\pi_c = \inf\{X_0 \mid X_T^\theta \geq c \ P\text{ -a.s. for some admissible } \theta\}.$$

The price is the least amount of initial capital needed to construct a self-financing wealth process whose terminal wealth exceeds the payoff of the claim almost surely.

Note that π_c is defined as a convex optimization problem over the set of admissible portfolio strategies and initial capitals X_0. It is an infinite dimensional linear optimization problem and, in principle, hard to solve. The following result can be seen as an application of "Lagrange multiplier method" from convex optimization, but we do not go into further details here.

Theorem 4.1. *Let Q be the equivalent martingale measure of the discounted price process $\tilde{S}$. If $E^Q|c|^2 < +\infty$, then*

$$\pi_c = e^{-rT} E^Q[c],$$

and there exists a self-financing wealth process $X^{\bar{\theta}}$ with admissible hedging strategy $\bar{\theta}$ and initial capital $X_0^{\bar{\theta}} = e^{-rT} E^Q c$ such that $X_T^{\bar{\theta}} = c$ almost surely. The $\bar{\theta}$ is given by $(\sigma^T)^{-1} z_t$ for z from the martingale representation theorem

$$\tilde{c} = E^Q \tilde{c} + \int_0^T z_t dB_t.$$

Proof. Let $X_0 \in \mathbb{R}$ and θ be admissible such that $X_T^\theta \geq c$ P-almost surely. Then $\tilde{X}_T^\theta \geq \tilde{c}$ P-almost surely. Since P and Q are equivalent, we also have $\tilde{X}_T^\theta \geq \tilde{c}$ Q-almost surely. Since $\theta\sigma \in \mathcal{H}^2_{\text{loc}}$, $\tilde{\theta}\sigma \in \mathcal{H}^2_{\text{loc}}$ and, by (3.7),

$$\tilde{X}_t^\theta = \tilde{X}_0 + \int_0^t \tilde{\theta}_s \sigma dB_s,$$

$\tilde{X}$ is a Q local martingale. Since θ is admissible, there is a Q-martingale M such that $\tilde{X}_t^\theta \geq M_t$ for all t. Let $(\tau^\nu)_{\nu=1}^\infty$ be a localizing sequence for $\tilde{X}^\theta$ so that each stopped process given by $\tilde{X}_{t\wedge\tau^\nu}^\theta$ is a true martingale and $\tilde{X}_{T\wedge\tau^\nu}^\theta \to \tilde{X}_T^\theta$. Since stopped processes are also bounded from below at $t = T$ by M_T which is Q-integrable, martingale property of the stopped processes and Fatou's lemma give

$$X_0^\theta = \tilde{X}_0^\theta = \liminf_\nu E^Q[\tilde{X}_{T\wedge\tau^\nu}^\theta] \geq E^Q[\tilde{X}_T^\theta] \geq E^Q[\tilde{c}] = e^{-rt} E^Q[c].$$

We have shown that

$$\pi_c \geq e^{-rt} E^Q[c].$$

To prove the other direction $\pi_c \leq e^{-rt} E^Q[c]$, we define a martingale $m_t := E^Q[\tilde{c} \mid \mathcal{F}_t]$. By the Martingale Representation Theorem 2.21, there exists $z \in \mathcal{H}^2$ such that

$$\tilde{c} = E^Q[\tilde{c}] + \int_0^T z_t dB_t.$$

Thus we have that $\tilde{X}_T^{\bar{\theta}} = \tilde{c}$ for $\tilde{X}_0 = E^Q[\tilde{c}]$ and for admissible $\bar{\theta}_t = (\sigma^T)^{-1} z_t$. Indeed, $\bar{\theta} \in \mathcal{H}^2_{\text{loc}}$ (actually, in $\mathcal{H}^2$),

$$\tilde{X}_t^{\bar{\theta}} = \tilde{X}_0 + \int_0^t \bar{\theta}_s \sigma dB_s,$$

and $\tilde{X}^{\bar{\theta}}$ is bounded from below by a Q-martingale, since it is a Q-martingale itself, by Theorem 2.18. Thus $\pi(c) \leq X_0^{\bar{\theta}} = \tilde{X}_0^{\bar{\theta}} = E^Q[\tilde{c}]$ and

$$\pi_c = E^Q[\tilde{c}].$$

The admissible $\bar{\theta}$ is the hedging strategy for c. $\square$

4.1 Delta-hedging of Vanilla options

In this section we consider Vanilla options

$$c = g(S_T) = g(S_T^0, \ldots, S_T^J)$$

for some g with quadratic growth. The idea is to combine the martingale characterization from Theorem 4.1 with Itô's formula to find a more explicit expression for the optimal hedging strategy.

We denote by $S = S^{t,x}$ the stochastic process describing the asset prices with "initial prices $x = (x^1, \ldots, x^d)$ at time t". Note that S does not have independent increments, but it is still a "Markov process" in the sense that its evolution depends on the past only through its current state. Most formulas of the previous sections can be written by replacing each initial prices by x^i and the initial time 0 by t. For instance, (3.5) reads, for $u \geq t$, as

$$(S^{t,x})_u^i = x^i \exp\left(\left(r - \frac{1}{2}\sum_{j=1}^d |\sigma^{ij}|^2\right)(u-t) + \sum_{j=1}^d \sigma^{ij}(B_u^j - B_t^j))\right), \tag{4.1}$$

where B is a Brownian motion under the martingale measure Q of $\tilde{S}$. It follows that, as in (4.3) $S^{t,x}$ solves the SDE (w.r.t B)

$$dS_u = \mathrm{diag}[S_u](rdu + \sigma dB_u) \quad u \in [t,T], \quad S_t = x. \tag{4.2}$$

In particular, S is an Itô process w.r.t. B, for $u \geq t$,

$$S_u^i = S_t^i + \int_t^u S_s^i(rds + \sum_{j=1}^d \sigma^{ij} dB_s^i) \tag{4.3}$$

so that we may apply Itô's formula w.r.t. B.

Since the option depends only on the terminal price of the underlying assets and S is Markov, we may define its (superhedging) price at time t by

$$\pi_c(t,x) = \inf\{\alpha \mid \alpha + \int_t^T \tilde{\theta}_u \sigma dB_u \geq g(S_T^{t,x}) \ P\text{ -a.s. for some admissible } \theta\}.$$

where we could assume "admissibility only on the interval $[t,T]$". The pricing formula of Theorem 4.1 becomes

$$\pi_c(t,x) = e^{-r(T-t)} E^Q[g(S_T^{t,x})|\mathcal{F}_t]. \tag{4.4}$$

In the next theorem, the optimal $\bar{z} = \partial_x \pi_c$ is called the *Delta-hedge*.

Theorem 4.2. *Assume that g has quadratic growth, $c = g(S_t)$, and that $\pi_c \in C^{1,2}$. Then the optimal (super-)hedging strategy is given by*

$$\bar{\theta}_t = \operatorname{diag}[S_t]\partial_x \pi_c(t, S_t) \quad t \in [0, T).$$

In particular, the amount $\bar{z}$ of assets in the optimal portfolios satisfies

$$\bar{z}^i_t = \partial_{x^i} \pi_c(t, S_t) \quad t \in [0, T).$$

Proof. Again, for simplicity, we assume in the proof that $d = 1$. By Theorem 4.1, the optimal solution is obtained from the stochastic integral representation of $Y_t := E^Q[\tilde{c}|\mathcal{F}_t]$ w.r.t B. By the Markov property of S, $E^Q[\tilde{c}|\mathcal{F}_t] = E^Q[\tilde{c}|S_t]$. Combining with (4.4), we have

$$Y_t := e^{-rt}\pi_c(t, S_t).$$

Applying Itô's formula to Y, we get

$$\begin{aligned} Y_t = {} & \pi_c(0, S_0) \\ & + \int_0^t (-re^{-rs}\pi_c(s, S_s) + e^{-rt}\partial_t \pi(s, S_s) + e^{-rs}\frac{1}{2}S_s^2\sigma^2\partial_{xx}\pi(s, S_s)ds \\ & + \int_0^t e^{-rs}\partial_x \pi(s, S_s)dS_s. \end{aligned}$$

Noting $dS_t = S_t(rdt + \sigma dB_t)$, this can be written as

$$\begin{aligned} Y_t = {} & \pi_c(0, S_0) \\ & + \int_0^t e^{-rs}\left(\partial_t \pi_c(s, S_s) + rS_s\partial_x \pi(s, S_s) + \frac{1}{2}S_s^2\sigma^2\partial_{xx}\pi(s, S_s) - r\pi_c(s, S_s)\right)ds \\ & + \int_0^t \tilde{S}_s \partial_x \pi(s, S_s)\sigma dB_s. \end{aligned}$$

Since Y is a Q-martingale and the last summand is a Q-martingale, the integral in the middle has to be zero for every t (a continuous martingale with a finite variation is a constant, a fact the we have not proved in these notes), so $\bar{\theta}_t = S_t\partial_x \pi(t, S_t)$ is the optimal portfolio. □

Remark 4.3. *We saw in the proof Theorem 4.2 that π_c has to solve the* Black Scholes partial differential equation *(PDE)*

$$\partial_t \pi_c + rx\partial_x \pi_c + \frac{1}{2}x^2\sigma^2\partial_{xx}\pi_c - r\pi_c = 0, \quad \pi_c(T, x) = g(x). \tag{4.5}$$

This draws a connection between PDE-theory and financial problems and provides one method for finding prices for Vanilla options. We will return to this later on at the end of the course.

4.2 Exercises

Exercise 4.2.1. *Write a Matlab function (as a m-file) that creates a sample path of a prices process in the 1-dimensional Black Scholes model. Write it as a function of initial time t with price x, terminal time T, $N+1$ equi-distant discretization points, drift μ, volatility σ and a sample of i.i.d standard normals. Plot some sample paths.*

Exercise 4.2.2. *Write a Matlab function (as a m-file) that creates a sample path of a price process in the 2-dimensional Black Scholes model. Write it as a function of initial time t with price vector x, terminal time T, $N+1$ equi-distant discretization points, drift vector μ, volatility matrix σ a sample of i.i.d standard normals vectors. Plot some sample paths (each coordinate in the same figure).*

Exercise 4.2.3. *Consider a European call option with strike K in the 1-dimensional Black Scholes model with $r=0$, $\sigma=1$ and $T=1$. Approximate the Delta hedge Δ by its piecewise constant approximation so that the resulting wealth process satisfies*

$$\Delta X_{t_k} = \Delta(t_{k-1}, S_{t_{k-1}})\Delta S_{t_k} \quad 0 = t_0 < \cdots < t_n = 1, X_0 = \pi_c(0, S_0);$$

see (3.2). *Simulate the wealth process. Based on this, does the discretized Delta hedge actually hedge the call option? Do this by plotting the differences between the simulated terminal wealths and the payoffs of the European call option (simulated "profit-losses").*

For this exercise, you need formulas that are proved in the following sections. These are the pricing functional π_c of the European call option (4.6) *and the Delta hedge*

$$\Delta(t,x) := \Phi(d_1(t,x)),$$

where Φ is the standard normal cumulative distribution function and

$$d_1(t,x) := \frac{\ln(x/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}.$$

4.3 Exercises

Exercise 4.3.1 (Integration by parts formula for Itô processes)**.** *Let W be a one dimensional Brownian motion and let X^i, $i=1,2$ be one dimensional Itô processes,*

$$X_t^i := X_0^i + \int_0^t \mu_s^i ds + \int_0^t \sigma_s^i dW_s \quad i = 1,2,$$

where, for $i=1,2$, $\mu^i \in \mathcal{H}^1_{loc}$ and $\sigma^i \in \mathcal{H}^2_{loc}$. Apply Itô's formula to $f(t,(x^1,x^2)) = x^1x^2$ to show the integration by parts formula

$$X_t^1 X_t^2 = X_0^1 X_0^2 + \int_0^t X_s^2 dX_s^1 + \int_0^t X_s^1 dX_s^2 + \int_0^t \sigma_s^1 \sigma_s^2 ds.$$

Exercise 4.3.2. *Let W be a d-dimensional Brownian motion, $b \in \mathbb{R}$ and $\sigma \in \mathbb{R}^d$, and consider the process*

$$S_t = \exp\{bt + \sigma \cdot W_t\}.$$

For $f \in C^{1,2}([0,T] \times \mathbb{R})$, show that $f(t, S_t)$ is an Itô process. Can you find a function f such that the process given by $Y_t := f(t, S_t)$ is a local martingale?

Exercise 4.3.3. *Prove that the discounted price process $\tilde{S}_t$ satisfies the SDE*

$$d\tilde{S}_t = \mathrm{diag}[\tilde{S}_t]\sigma(\lambda dt + dW_t).$$

Exercise 4.3.4. *Find the pricing function $\pi(t,x)$ and the optimal hedging strategy for the quadratic claim $c = g(S_T) := S_T^2$. Verify your solution π by checking that it solves the Black Scholes partial differential equation* (4.5).
Hint: Recall that $E^Q e^{a\eta} = e^{\frac{1}{2}a^2}$ for a standard normally distributed η under Q.

4.4 The Black-Scholes formula for puts and calls

In the one dimensional case, the price process can be written (see (3.5))

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

where B is a Brownian motion under the martingale measure Q of the discounted price process $\tilde{S}$.

Theorem 4.4. *The superhedging price of be the European call option $c := (S_T - K)^+$ written at time t with the strike K and maturity T is*

$$\pi_c(t, S_t) = S_t \Phi(d_1) - \Phi(d_2) K e^{-r(T-t)}, \tag{4.6}$$

where Φ is the cumulative distribution function of the standard normal distribution and

$$d_1 = \frac{\ln(S_t/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$
$$d_2 = \frac{\ln(S_t/K) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}.$$

Proof. Let $x = S_t$ and $S_T^{t,x} = x \exp\left(\left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma(B_T - B_t)\right)$. By Theorem 4.1,

$$\pi_c(t, S_t) = e^{-r(T-t)} E^Q[c].$$

Denoting $\eta \sim N(0,1)$, we get

$$\begin{aligned} E^Q[(S_T^{t,x} - K)^+] &= E^Q[(\exp(\ln(S_T^{t,x}) - K)^+] \\ &= E^Q[\exp(\ln x + (r - \frac{1}{2}\sigma^2)(T-t) + \sigma\sqrt{T-t} \cdot \eta) - K)^+]. \end{aligned}$$

Denoting $z := \ln x + (r - \frac{1}{2}\sigma^2)(T-t)$ and $\phi(y) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ (the density of a standard normal distribution),

$$
\begin{aligned}
E^Q[(S_T^{t,x} - K)^+] & \int_{z+\sigma\sqrt{T-t}y > \ln K} (e^{z+\sigma\sqrt{T-t}y} - K)\phi(y)dy \\
&= \int_{\frac{\ln K - z}{\sigma\sqrt{T-t}}}^{\infty} e^{z+\sigma\sqrt{T-t}y}\phi(y)dy - K\int_{\frac{\ln K - z}{\sigma\sqrt{T-t}}}^{\infty} \phi(y)dy \\
&=: I - II.
\end{aligned}
$$

Using the symmetry of ϕ, the second term can be written as

$$
II = K\int_{\frac{\ln K - z}{\sigma\sqrt{T-t}}}^{\infty} \phi(y)dy = K\Phi\left(\frac{z - \ln K}{\sigma\sqrt{T-t}}\right).
$$

As to the first term,

$$
\begin{aligned}
I &= \int_{\frac{\ln K - z}{\sigma\sqrt{T-t}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{z+\sigma\sqrt{T-t}y - \frac{1}{2}y^2} dy \\
&= xe^{r(T-t)} \int_{\frac{\ln K - z}{\sigma\sqrt{T-t}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\sigma\sqrt{T-t})^2} dy \\
&= xe^{r(T-t)} \int_{\frac{\ln K - z}{\sigma\sqrt{T-t}} - \sigma\sqrt{T-t}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\tilde{y}^2} d\tilde{y} \\
&= xe^{r(T-t)}\Phi\left(\sigma\sqrt{T-t} - \frac{\ln K - z}{\sigma\sqrt{T-t}}\right),
\end{aligned}
$$

where the last line follows from the symmetry of the ϕ. Combining, we get

$$
\pi(c_C) = x\Phi\left(\sigma\sqrt{T-t} - \frac{\ln K - z}{\sigma\sqrt{T-t}}\right) - \Phi\left(\frac{z - \ln K}{\sigma\sqrt{T-t}}\right)Ke^{-r(T-t)}.
$$

Substituting z and simplifying gives the result. □

4.4.1 Put-call parity

Given a common strike price K and maturity T, the payoffs c_P and c_C of the European put and European call satisfy

$$
c_P = (K - S_T)^+ = K - S_T + (S_T - K)^+ = c_C - (S_T - K).
$$

Here we may identify $(S_T - K)$ with the payoff

$$
c_F := S_T - K
$$

of a *forward contract*. The price π_F of c_F can simply be calculated from (4.4),

$$
\pi_F(t, x) = e^{-r(T-t)} E^Q[S_T^{t,x} - K | \mathcal{F}_t] = x - e^{-r(T-t)}K,
$$

where we used the fact that Q is the martingale measure of $\tilde{S}$. Since the pricing functional is linear (in a "complete model" like the Black Scholes model), we get the *put-call parity*

$$\pi_P(t,x) = \pi_C(t,x) - x + e^{-r(T-t)}K. \tag{4.7}$$

The put-call parity in conjunction with the Black Scholes formula of the call option in Theorem 4.4 gives the Black Scholes formula for the put option.

Theorem 4.5. *The superhedging price of be the European put option $c := (K - S_T)^+$ written at time t with the strike K and maturity T is*

$$\pi_P(t,S_t) = -S_t\Phi(-d_1) + \Phi(-d_2)Ke^{-r(T-t)},$$

where Φ is the cumulative distribution function of the standard normal distribution and

$$d_1 = \frac{\ln(S_t/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$
$$d_2 = \frac{\ln(S_t/K) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}.$$

Proof. Combine Theorem 4.4, (4.7) and use the fact that $\Phi - \frac{1}{2}$ is antisymmetric. □

4.5 Sensitivity analysis and Greeks

The Greeks refer to sensitivity of the derivative prices with respect to the model parameters and underlying prices. Throughout this section, π is the price of some given claim written at time t with current underlying prices $S_t = x$.

The most important Greek is the Delta,

$$\Delta := \partial_x,$$

since it gives the sensitivity with respect to the underlying prices. By Theorem 4.2, Delta provides a formula for the optimal hedging portfolios of Vanilla options.

Example 4.6. *Recall that, for a European call option $c = (S_T - K)^+$,*

$$\pi_c(t,x) = x\Phi(d_1) - \Phi(d_2)Ke^{-r(T-t)},$$

where Φ is the cdf of the standard normal distribution,

$$d_1 = \frac{\ln(x/K) + r(T-t) + \frac{1}{2}v^2}{v}$$
$$d_2 = \frac{\ln(x/K) + r(T-t) - \frac{1}{2}v^2}{v}$$

and $v = \sigma\sqrt{T-t}$. The Delta becomes

$$\Delta := \Phi(d_1).$$

In particular, by Theorem 4.2, the optimal hedging portfolio consists of holding

$$\bar{z}_t = \Phi\left(\frac{\ln(S_t/K) + r(T-t) + \frac{1}{2}v^2}{v}\right)$$

assets at every $t \in [0, T)$.

Proof. Denoting $\phi(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}$,

$$\Delta = \partial_x \pi_c(t,x) = \Phi(d_1) + \phi(d_1)\frac{K}{v} - \phi(d_2)Ke^{-r(T-t)}\frac{K}{vx}.$$

Since $d_2 = d_1 - v$ and $d_1 v - \frac{1}{2}v^2 = \ln(x/K) + r(T-t)$, we have

$$\phi(d_2) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(d_1-v)^2} = \phi(d_1)e^{d_1 v - \frac{1}{2}v^2} = \phi(d_1)e^{r(T-t)}\frac{x}{K},$$

so $\Delta = \Phi(d_1)$. □

In practice, a closed form solution of $\pi_c(t,x)$ is not available for general Vanilla option c, so one has to resort to numerical methods when constructing optimal hedging strategies. This will be a topic of the next section.

The Greek *gamma* of an option is the second derivative with respect to the underlying,

$$\Gamma := \partial_{xx}\pi.$$

For Delta-hedges, it tells how quickly the hedging portfolio has to be adjusted to the price movements of the underlying. The "Greek" Vega, is another important quantity defined as

$$\mathcal{V} := \partial_\sigma \pi.$$

This gives the sensitivity with respect to the volatility. Since the volatility is a parameter of the model, a high absolute value in Vega indicates higher model risk. Practitioners often seek to build a portfolio with low Vega in order to decrease the risk of a miss-specified model.

The Greek *Theta* is the rate of change with respect to time,

$$\Theta = \partial_t \pi,$$

also called the *time value of the option.* Sometimes, it defined as the derivative w.r.t. to the terminal time T. Finally, Rho is the sensitivity to the instantaneous interest rate.

$$\rho := \partial_r \pi.$$

In practice, Black Scholes model is usually extended so that it takes into account the whole term structure of interest rates ("time dependent" $r(t)$). On the other hand, interest rates vary slowly if at all before the maturity of the option, so the role of Rho is less relevant in practice.

Exercise 4.5.1. *Show that, for the European call option,*

$$\Gamma = \frac{\phi(d_1)}{\sigma x\sqrt{T-t}}.$$

Exercise 4.5.2. *Show that, for the European call option,*

$$\mathcal{V} = x\sqrt{T-t}\phi(d_1).$$

Exercise 4.5.3. *Show that, for the European call option,*

$$\Theta = -\frac{\sigma x\phi(d_1)}{2\sqrt{T-t}} - rKe^{-r(T-t)}\Phi(d_2).$$

Exercise 4.5.4. *Show that, for the European call option,*

$$\rho = K(T-t)e^{-r(T-t)}\Phi(d_2).$$

5 Monte Carlo methods in pricing

We are interested in computing prices of derivatives c, which by Theorem 4.1, boils down to computing

$$e^{-r(T-t)}E^Q[c]$$

as soon as $E^Q|c|^2 < \infty$. The basis of Monte Carlo methods is the law of large numbers.

Let $(\xi^{(m)})_{m=1}^{\infty}$ be a sequence independent and identically distributed random variables with $\xi^{(n)} \sim \eta$ for $E|\xi| < \infty$.

Theorem 5.1 (Strong law of large numbers)**.** *We have, almost surely,*

$$\frac{1}{M}\sum_{m=1}^{M}\xi^{(m)} \to E\eta.$$

The random variable

$$\mu^{(M)} = \frac{1}{M}\sum_{m=1}^{M}\xi^{(m)},$$

called *the sample mean*, is used as an estimate of the expectation $E\xi$. This random variable, a particular "estimator of $E\xi$", satisfies

$$E\mu^{(M)} = E\xi$$
$$\operatorname{std}\left(\mu^{(M)}\right) = \frac{1}{\sqrt{M}}\operatorname{std}(\xi).$$

The first equation tells us that the sample mean is "unbiased" and the second tells us that the "rate of convergence is of order $\sqrt{M}$". That is, if we want to reduce the standard deviation of the estimator to one tenth, we need to use hundred times larger "sample size M".

Notice also that the central limit theorem, Theorem 2.12, tells us that, when $\mathrm{std}(\eta)$ is finite,

$$\sqrt{M}(\mu^{(M)} - E\eta) \xrightarrow{d} \xi \sim N(0, \mathrm{std}(\xi)^2).$$

Thus the estimator is asymptotically normal. Loosely speaking, the estimator is "approximately normally distributed for large M around $E\xi$",

$$\frac{1}{M}(\sum_{m=1}^{M}(\xi^{(m)} - E\xi)) = \frac{1}{\sqrt{M}}\frac{1}{\sqrt{M}}(\sum_{m=1}^{M}(\xi^{(m)} - E\xi)) \sim\sim N\left(0, \left(\frac{\mathrm{std}(\xi)}{\sqrt{M}}\right)^2\right).$$

To derive confidence intervals for the estimator $\mu^{(M)}$, we introduce the "sample variance"

$$(\sigma^{(M)})^2 := \frac{1}{M-1}\sum_{m=1}^{M}\left(\xi^{(m)} - \mu^{(M)}\right)^2$$

to estimate $\mathrm{Var}(\xi)$. We have

$$\begin{aligned}
(\sigma^{(M)})^2 &= \frac{M}{M-1}\sum_{m=1}^{M}\left(\frac{(\xi^{(m)})^2}{M} - 2\frac{\xi^{(m)}}{M}\mu^{(M)} + \frac{(\mu^{(M)})^2}{M}\right) \\
&= \frac{M}{M-1}\left(\sum_{m=1}^{M}\frac{(\xi^{(m)})^2}{M} - (\mu^{(M)})^2\right)
\end{aligned}$$

so, by the law of large numbers,

$$(\sigma^{(M)})^2 \to \left(E(\xi)^2 - (E\eta)^2\right) = \mathrm{Var}(\xi).$$

We leave it as an exercise to verify that $(\sigma^{(M)})^2$ is also an unbiased estimator of $\mathrm{Var}(\xi)$ (this is an elementary but a bit lengthier computation).

In the exercises, we use "heuristic confidence intervals" by approximating

$$\mu^{(M)} \sim\sim N(E\xi, \frac{(\sigma^{(M)})^2}{M})$$

so that, "with large M, we have with 95% probability" that $E\eta$ is in the interval

$$[\mu^{(M)} - \frac{1.96\sigma^{(M)}}{\sqrt{M}}, \mu^{(M)} + \frac{1.96\sigma^{(M)}}{\sqrt{M}}]. \tag{5.1}$$

Here we put aside the fact that $(\sigma^{(M)})^2$ is just an estimator of $\mathrm{Var}(\xi)$. It would be more justified to do asymptotic analysis for the estimator $(\sigma^{(M)})^2$ and derive confidence intervals for it, which would increase the confidence intervals of $\mu^{(M)}$.

5.1 MC of Vanilla options

Consider a Vanilla option $c = f(S_T)$ such that $E^Q|f(S_T)|^2 < \infty$. Taking independent copies $(S_T^{t,x})^{(m)}$, $m = 1, 2, \ldots$ of $S_T^{t,x}$, the law of large numbers tells us that, almost surely,

$$\frac{1}{M}\sum_{m=1}^{M} f(S_T^{(m)}) \to E^Q[f(S_T^{t,x})].$$

Since

$$S_T^{t,x} \sim x \exp\left(\left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma\sqrt{T-t}\eta\right),$$

where η has a normal standard distribution under Q, we know how to generate independent copies of $S_T^{t,x}$ under Q, and we get the following Monte Carlo method.

Algorithm 5.2. *An MC algorithm (with confidence intervals) for prices of Vanilla options $c = f(S_T)$ with initial time t and price x.*

1. *Generate a mutually independent sample*

$$\{\eta^{(1)}, \ldots, \eta^{(M)}\}$$

from a standard normal distribution.

2. *Compute, for $m = 1, \ldots, M$*

$$\xi^{(m)} = e^{-r(T-t)} f\left(xe^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma\sqrt{T-t}\cdot\eta^{(m)}}\right).$$

3. *Compute the MC estimate*

$$\mu^{(M)} = \frac{1}{M}\sum_{m=1}^{M}\xi^{(m)}.$$

4. *Compute the confidence intervals according to* (5.1).

Note that in the above algorithm, there is no need to simulate the whole path of the price process, which saves a lot of computational costs.

5.2 MC and Delta-hedging of Vanilla options

We continue with a Vanilla option $c = f(S_T)$ such that $E^Q|f(S_T)|^2 < \infty$ and that f has quadratic growth. Recall from Theorem 4.2 that the optimal hedging strategy $\bar{z}$ is given by

$$\partial_x \pi_c(t, x).$$

Assuming that, below, we may change the order of integration and differentiation, we have

$$\begin{aligned}\partial_x \pi_c(t,x) &= e^{-r(T-t)} \partial_x E^Q[f(S_T^{t,x})] \\ &= e^{-r(T-t)} E^Q[\partial f(S_T^{t,x}) \partial_x S_T^{t,x}] \\ &= e^{-r(T-t)} E^Q[\partial f(S_T^{t,x}) S_T^{t,1}],\end{aligned}$$

where the last line follows from the $S_T^{t,x} = x \exp\left(\left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma(B_T - B_t)\right)$. Moreover, it is sufficient that f is merely "weakly differentiable", so ∂f is understood in this sense.

Algorithm 5.3. *An MC algorithm for Delta-hedge of a Vanilla option $c = f(S_T)$ with initial time t and price x.*

1. *Generate a mutually independent sample*

$$\{\eta^{(1)}, \dots, \eta^{(M)}\}$$

from a standard normal distribution.

2. *Compute*

$$e^{-r(T-t)} \frac{1}{M} \sum_{m=1}^{M} \left[f'\left(x e^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma\sqrt{T-t}\cdot\eta^{(m)}} \right) e^{(r-\frac{1}{2}\sigma^2)(T-t)+\sigma\sqrt{T-t}\cdot\eta^{(m)}} \right].$$

Of course, in step 2 above, it is possible to cancel out the terms $e^{-r(T-t)}$, depending on how the computation is actually implemented.

Example 5.4. *For the European call option, the "weak derivative" of the payoff $f(x) = (x-K)^+$ becomes*

$$f'(x) = \begin{cases} 0 & \text{if } x \le K \\ 1 & \text{if } x > K. \end{cases}$$

Another method to estimate Delta is to approximate

$$\partial_x \pi(t,x) = \lim_{h \to 0} \frac{\pi(t, x+h) - \pi(t,x)}{h}$$

by difference quotients with small h:

$$\partial_x \pi(t,x) \approx \frac{\pi(t, x+h) - \pi(t,x)}{h}.$$

For this, we may apply Algorithm 5.2 to $\pi(t, x+h)$ with the given x and "small" h (with the same sample for h). This is easier to implement, especially when ∂f is not available, but is more susceptible to simulation errors. Both these methods can be modified to get MC-algorithms for estimating other Greeks as well.

5.3 Exercises

Exercise 5.3.1. *Write Matlab functions (as m-files) for the pricing functional* π *of the European call option and its Greeks. Write them as functions of all the parameters,* t, x, T, r, σ, *and* K. *Plot each Greek (choose three) as a function of the corresponding parameter,* Δ *as a function of* x, $\mathcal{V}$ *as a function of* σ, *etc.*

Exercise 5.3.2. *Apply MC algorithm to estimate prices for the European call option. Plot the estimated prices as a function of the number* M *of simulations. Plot also the confidence intervals according to* (5.1). *After finding* M *large enough, estimate the prices as a function of* x, *for fixed* $t = 0$, T, r, σ *and* K. *Plot the estimates as a function of* x *in the same figure with correct prices from Theorem 4.4.*

Exercise 5.3.3. *Apply MC algorithm to estimate Greeks* Δ, Γ *and* $\mathcal{V}$ *for the European call option. Plot the estimates in the same figure with the theoretical values. Plot them as functions of the corresponding and one additional parameter, e.g.,* Δ *as a function of* x *and* σ, $\mathcal{V}$ *as a function of* σ *and* r, *etc.*

5.4 MC for Barriers and Asians

Consider now a path-dependent claim c such that $E^Q|c|^2 < \infty$. The first step is to discretize time into N intervals with a time grid $\{t_0, \ldots, t_N\}$, $t_0 = 0$, $t_N = T$, so that we may write the claim as $c = f(S_{t_0}, \ldots, S_{t_N})$. In fact, this is how we wrote the path-dependent payoff functions in Exercises 2.2. After that, MC algorithm is built on the whole sample paths of the price process. Since

$$S_{t_n} = S_{t_{n-1}} \exp\left(\left(r - \frac{1}{2}\sigma^2\right)(t_n - t_{n-1}) + \sigma\sqrt{t_n - t_{n-1}}\,\eta^{(n)})\right), \tag{5.2}$$

where $(\eta^{(n)})_{n=1}^N$ are i.i.d., and standard normally distributed under Q, we get the following Monte Carlo method.

Algorithm 5.5. *An MC algorithm for the price of a path-dependent option*

$$c = f(S_{t_0}, \ldots, S_{t_N}).$$

1. *For* $m = 1, \ldots, M$,
   (a) *Generate a mutually independent sample*
   $$\{\eta^{(1)}, \ldots, \eta^{(N)}\}$$
   from a standard normal distribution.
   (b) *For* $n = 1, \ldots, N$, *compute*
   $$s_n = s_{n-1} e^{\left(r - \frac{1}{2}\sigma^2\right)(t_n - t_{n-1}) + \sigma\sqrt{t_n - t_{n-1}}\,\eta^{(n)})},$$
   where $s_0 = S_0$.

(c) *Compute* $f_m = f(s_0, \dots, s_N)$

2. *Compute*

$$e^{-rT}\frac{1}{M}\sum_{m=1}^{M} f_m.$$

In the actual implementation, it possible to parallelize step 1. On the other hand in Matlab, it is possible to simulate M different paths simultaneously and do the computations in steps (b) and (c) in the vector form.

The general MC algorithm described above can be tuned to the option at hand. For example, for Barrier options, we can do the following:

Algorithm 5.6. *An MC algorithm for the price of the down-and-out Barrier option*

$$c = (S_T - K)^+ \mathbb{1}_{\mathbb{R}_+}(\min_n S_{t_n} - B).$$

1. *For* $m = 1, \dots, M$*,*
   (a) *For* $n = 1, \dots, N$
      i. *Generate* $\eta^{(n)}$ *from a a standard normal distribution*
      ii. *Compute*

$$s_n = s_{n-1} e^{\left(r - \frac{1}{2}\sigma^2\right)(t_n - t_{n-1}) + \sigma\sqrt{t_n - t_{n-1}}\eta^{(n)})}.$$

      iii. *If* $s_n < B$*, define* $f_m = 0$ *and go to 1.*
   (b) *Set* $f_m = (s_N - K)^+$*.*

2. *Compute*

$$e^{-rT}\frac{1}{M}\sum_{m=1}^{M} f_m.$$

6 The Black Scholes PDE and finite difference method

We saw in the proof Theorem 4.2 that, for the Vanilla option $c = f(S_T)$ with appropriate growth conditions, the corresponding pricing function $\pi : [0,T] \times \mathbb{R}_+ \mapsto \mathbb{R}$ solves the Black Scholes partial differential equation (BS-PDE)

$$\partial_t \pi + rx\partial_x \pi + \frac{1}{2}x^2\sigma^2\partial_{xx}\pi - r\pi = 0, \quad \pi(T,x) = f(x). \tag{6.1}$$

Defining the parameters

$$a = \frac{r}{\sigma^2} - \frac{1}{2}, \quad b = \frac{1}{4}\left(\frac{2r}{\sigma^2} + 1\right)^2$$

and a function

$$u(\tau, y) = e^{ay+b\tau}\pi(T - 2\tau/\sigma^2, e^y) \tag{6.2}$$

on $[0, \frac{1}{2}\sigma^2 T] \times \mathbb{R}$, we have (below, we omit the arguments (τ, y) for u and $(T - 2\tau/\sigma^2, e^y)$ from π and its derivatives)

$$\begin{aligned}
\partial_\tau u &= bu - \frac{2}{\sigma^2} e^{ay+b\tau}\partial_t \pi \\
\partial_y u &= au + e^{(a+1)y+b\tau}\partial_x \pi \\
\partial_{yy} u &= a^2 u + ae^{(a+1)y+b\tau}\partial_x \pi + (a+1)e^{(a+1)y+b\tau}\partial_x \pi + e^{(a+2)y+b\tau}\partial_{xx}\pi.
\end{aligned}$$

Combining this with the fact that π solves (6.1), we get by a direct (but tedious) verification that u solves the *heat equation*

$$\partial_\tau u = \partial_{yy} u, \quad u(0, y) = e^{ay} g(e^y).$$

The formula (6.2) can be inverted so that

$$\pi(t, x) = e^{-a \ln x - b\frac{1}{2}\sigma^2(T-t)} u(\frac{1}{2}\sigma^2(T-t), \ln x). \tag{6.3}$$

Therefore, numerical solutions of u yield numerical solutions to π.

6.1 The finite difference method

We choose an evenly spaced time discretization $0 = \tau_0 \leq \dots \tau_N$ for a fixed integer N. We truncate the y-state space $\mathbb{R}$ to $[y_{min}, y_{max}]$, and choose an evenly spaced space discretization $y_{min} = y_0 \leq \dots, y_J = y_{max}$ for a fixed integer J.

Denoting $u_n^j = u(\tau_n, y_j)$, we approximate, for every n and $0 < j < J$ (the boundaries are dealt with later on)

$$\begin{aligned}
\partial_\tau u(\tau_n, y_j) &= \frac{u_{n+1}^j - u_n^j}{\Delta\tau_n} \\
\partial_y u(\tau_n, y_j) &= \frac{u_n^{j+1} - u_n^{j-1}}{2\Delta y_j} \\
\partial_{yy} u(\tau_n, y_j) &= \frac{u_n^{j+1} - 2u_n^j + u_n^{j-1}}{(\Delta y_j)^2}.
\end{aligned}$$

Here $\partial_\tau u$ is approximated by the "forward difference" and $\partial_{yy} u$ by the "second-order central difference". Using these finite differences, our approximation for the heat equation $\partial_\tau u = \partial_{yy} u$ becomes

$$\frac{u_{n+1}^j - u_n^j}{\Delta\tau_n} = \frac{u_n^{j+1} - 2u_n^j + u_n^{j-1}}{(\Delta y_j)^2}$$

which we write in the recursive form

$$u_{n+1}^j = \frac{\Delta\tau_n}{(\Delta y_j)^2} u_n^{j+1} + \left(1 - \frac{2\Delta\tau_n}{(\Delta y_j)^2}\right) u_n^j + \frac{\Delta\tau_n}{(\Delta y_j)^2} u_n^{j-1}. \tag{6.4}$$

This is an *explicit method* since we used forward differences in time. The method is numerically stable if $\Delta\tau_n \leq (\Delta y_n)^2/2$; a fact which we state without a proof. Other choices of derivative approximations lead to other finite difference methods (implicit method, Crank-Nicolson method, etc.).

The initial condition becomes (for $j = 0$ and $j = J$, these are given later on)

$$u_0^j = e^{a(j\Delta y_n + y_{min})} f(e^{j\Delta y_n + y_{min}}) \quad j = 1, \ldots, J-1.$$

The boundary conditions along the truncation boundary are less obvious and not given by the data. Here we choose these boundary conditions by asymptotic arguments.

Consider first the points $y_0^j = y_{min}$. Since

$$u(\tau, y) = e^{ay+b\tau}\pi(T - 2\tau/\sigma^2, e^y) = e^{ay}e^{(b-r2/\sigma^2)\tau}E^Q[f(S^{T-2\tau/\sigma^2, e^y})],$$

we get, if $f(0)$ is finite, that

$$\lim_{y\to-\infty} \frac{u(\tau, y)}{e^{ay}} = \frac{e^{ay}e^{(b-r2/\sigma^2)\tau}E^Q[f(S^{T-2\tau/\sigma^2,0})]}{e^{ay}} = e^{(b-r2/\sigma^2)\tau}f(0).$$

Thus it is natural to set

$$u_n^0 = e^{(b-2r/\sigma^2)n\Delta\tau_n + ay_{min}} f(0) \quad n = 0, \ldots, N.$$

Remark 6.1. *If $f(0)$ is not finite at the origin, it is possible to refine the argument using L'Hôpital's rule as below; we omit the details.*

Consider now the boundary points $y_N^j = y_{max}$. Assuming that we may differentiate under the integral sign, change the order of integration and the limit, and that f has a linear growth, we get from L'Hôpital's rule that

$$\begin{aligned}
\lim_{y\to-\infty} \frac{u(\tau, y)}{e^{ay}f(e^y)} &= \lim_{y\to\infty} \frac{e^{(b-r2/\sigma^2)\tau}E^Q[f(e^y S^{T-2\tau/\sigma^2,1})]}{f(e^y)} \\
&= e^{(b-r2/\sigma^2)\tau}E^Q[S_T^{T-2\tau/\sigma^2,1}] \\
&= e^{(b-r2/\sigma^2)\tau}e^{r(2\tau/\sigma^2)} \\
&= e^{b\tau}.
\end{aligned}$$

Thus it makes sense to approximate $u(\tau, y) \sim e^{b\tau}e^{ay}f(e^y)$ for large y. Thus we may set

$$u_n^J = e^{ay_{max} + bn\Delta\tau_n} f(e^{y_{max}}) \quad n = 0, \ldots, N.$$

Remark 6.2. *If f has quadratic growth, the above argument has to be modified.*

In this case, applying L'Hôpital's rule twice,

$$\begin{aligned}
\lim_{y\to-\infty}\frac{u(\tau,y)}{e^{ay}f(e^y)} &= \lim_{y\to\infty}\frac{e^{(b-r2/\sigma^2)\tau}E^Q[f(e^yS^{T-2\tau/\sigma^2,1})]}{f(e^y)} \\
&= e^{(b-r2/\sigma^2)\tau}E^Q[(S_T^{T-2\tau/\sigma^2,1})^2] \\
&= e^{(b-r2/\sigma^2)\tau}e^{(2r+\sigma^2)(2\tau/\sigma^2)} \\
&= e^{(b+r2/\sigma^2+2)\tau}.
\end{aligned}$$

Here the second last line follows from

$$E^Q[(S_T^{t,x})^2] = x^2e^{(2r+\sigma^2)(T-t)},$$

which was computed as a part of Exercise 4.3.4. We end up with the boundary conditions

$$u_n^J = e^{ay_{max}+(b+r2/\sigma^2+2)n\Delta\tau_n}f(e^{y_{max}}) \quad n = 0,\ldots,N.$$

Now we are ready to formulate an algorithm based on these observations.

Algorithm 6.3. *A finite difference method for approximating u corresponding to the price function of a Vanilla option $c = f(S_T)$ with linearly growing f that is finite at the origin.*

1. *Choose N, J and y_{max} so that $\Delta\tau := \frac{1}{2}\sigma^2T/N$ an $\Delta y := (y_{max}-y_{min})/J$ satisfy $\Delta\tau_n \leq (\Delta y_n)^2/2$.*
2. *For $n = 0,\ldots,N$, set*

$$\begin{aligned}
u_n^0 &= e^{(b-r2/\sigma^2)n\Delta\tau_n+ay_{min}}f(0) \\
u_n^J &= e^{ay_{max}+bn\Delta\tau_n}f(e^{y_{max}}).
\end{aligned}$$

3. *For $j = 1,\ldots,J-1$, set*

$$u_0^j = e^{a(j\Delta y_n+y_{min})}f(e^{j\Delta y_n+y_{min}}) \quad j = 1,\ldots,J-1.$$

4. *For $n = 1,\ldots,N$, $j = 1,\ldots,J-1$, compute*

$$u_{n+1}^j = \frac{\Delta\tau_n}{(\Delta y_n)^2}u_n^{j+1} + \left(1-\frac{2\Delta\tau_n}{(\Delta y_n)^2}\right)u_n^j + \frac{\Delta\tau_n}{(\Delta y_n)^2}u_n^{j-1}.$$

In Matlab, the steps are most convenient to implement using vectors, in particular, in the recursion step 4, by introducing a tridiagonal matrix.

6.2 Exercises

Exercise 6.2.1. *Find suitable discretization parameters N, J and the cutoff values y_{max} and y_{min} so that the finite difference method approximates the theoretical price of a European call option with 0.1 % percent accuracy for some fixed (and nonzero) parameters.*

Exercise 6.2.2. *Apply the finite difference method to approximate the prices of an European put option and the quadratic option $c = (S_T)^2$.*

Exercise 6.2.3. *Estimate the Greek Θ on the whole interval $[0, T]$ of a European call option using the finite difference method.*

Exercise 6.2.4. *Estimate the Greek Γ at time $t = 0$ and at some σ of a European call option using the finite difference method.*

7 Variance reduction methods in Monte Carlo

Let Z be a random variable and assume that $E^Q[G(Z)^2] < \infty$ for a function G with quadratic growth. The "naive Monte Carlo simulation" under Q gives estimates of

$$\alpha := E^Q[G(Z)]$$

that are proportional to the variance of $G(Z)$ under Q. The general idea of variance reduction methods is to seek another random variable Y and possibly another probability measure Q_h such that

$$E^{Q_h}[Y] = E^Q[G(Z)]$$

and such that the variance of Y of under Q_h is smaller. Then performing Monte Carlo for Y under Q_h results in smaller confidence intervals.

7.1 Importance sampling

The idea in importance sampling is to look for a measure Q_h under which the variance of the Monte Carlo estimator is smaller. Minimizing variance amounts to minimizing the second moment.

Consider Z whose law has the density g w.r.t. the Lebesgue measure under Q and another density function h such that $h > 0$ on $\{g > 0\}$. We have

$$E^Q[G(Z)] = \int G(x)g(x)dx = \int G(x)\frac{g(x)}{h(x)}h(x)dx = E^h\left[G(Z)\frac{g(Z)}{h(Z)}\right],$$

where E^h denotes the expectation w.r.t. Q_h under which Z has the density h [Note that $dQ/dQ_h = g(Z)/h(Z)$ if $\{h > 0\} = \{g > 0\}$]. In particular, the sample mean (Monte Carlo estimator)

$$\frac{1}{M}\sum_{m=1}^{M} G(Z^{(m)})\frac{g(Z^{(m)})}{h(Z^{(m)})}$$

is an unbiased estimator of α (here $Z^{(m)}$ are distributed according to the density). The second moment of $G(Z)\frac{g(Z)}{h(Z)}$ under Q_h is

$$\begin{aligned} E^h\left[G(Z)\frac{g(Z)}{h(Z)}\right]^2 &= \int \left[G(x)\frac{g(x)}{h(x)}\right]^2 h(x)dx \\ &= E^Q[G(Z)^2\frac{g(Z)}{h(Z)}]. \end{aligned}$$

If $G > 0$ on $\{g > 0\}$, the minimal second moment (and thus minimal variance) is obtained with the choice $h(x) = \frac{G(x)g(x)}{\alpha}$, in which case the variance is zero. However, α is what we are trying to estimate, so this choice is not available to us.

Importance sampling by the change of drift

Assume now that g is the multivariate density distribution on $\mathbb{R}^n$ with mean vector 0 covariance matrix $\mathbb{1}_n$. In our pricing applications, all sampling is based on this structure. For instance, for a path-dependent option we have

$$G(Z^1, \dots Z^n) = f(S_1, \dots, S_n),$$

where $S_i = S_{i-1}e^{(r-\frac{1}{2}\sigma^2)\delta t+\sigma\sqrt{\Delta t}Z_i}$. Recall that the density (w.r.t. the Lebesgue measure) $h(\cdot;\mu,\Sigma)$ of a n-dimensional normal distribution with mean μ and covariance matrix Σ is

$$h(x;\mu,\sigma) = \frac{1}{\sqrt{(2\pi)^n \det\Sigma}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Let h be the multivariate density distribution on $\mathbb{R}^n$ with drift vector μ and covariance matrix $\mathbb{1}_n$. Under Q_h, we have $Z \sim N(\mu, \mathbb{1}_n)$. Thus, under Q_h, Z equals in distribution to $\tilde{Z}+\mu$, where $\tilde{Z} \sim N(0, \mathbb{1}_n)$. Since $Z \sim N(0, \mathbb{1}_n)$ under Q, this means that

$$E^h\left[G(Z)\frac{g(Z)}{h(Z)}\right] = E^Q\left[G(Z+\mu)\frac{g(Z+\mu)}{h(Z+\mu)}\right],$$

so, in this setting, performing Monte Carlo with importance sampling amounts to performing the Monte Carlo under the original measure Q for the random variable

$$G(Z+\mu)\frac{g(Z+\mu)}{h(Z+\mu)},$$

where $Z \sim N(0, \mathbb{1}_n)$. A direct computation gives

$$\frac{g(Z)}{h(Z)} = e^{-Z^T\mu + \frac{1}{2}\mu^T\mu}.$$

The second moment can be written as

$$E^h\left[G(Z)\frac{g(Z)}{h(Z)}\right]^2 = E^Q[G(Z)^2 e^{-Z^T\mu+\frac{1}{2}\mu^T\mu}],$$

where the right hand side is a convex function of μ. Applying Monte Carlo (in the context of optimization, also known as sample average approximation), we approximate the right side by

$$\frac{1}{\tilde{M}}\sum_{m=1}^{\tilde{M}}\left(G(Z^{(m)})^2 e^{-(Z^{(m)})^T\mu+\frac{1}{2}\mu^T\mu}\right),$$

where $Z^{(1)}, \dots Z^{(\tilde{M})}$ is a mutually independent sample of Z. Minimizing this over μ is a finite dimensional convex optimization problem.

Remark 7.1. *It is also possible to vary the covariance matrix* Σ *and not just the drift* μ *and still obtain a convex optimization problem. This involves a change of variables in terms of the* natural parameters

$$\theta := \begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix} := \begin{pmatrix}\Sigma^{-1}\mu\\ -\frac{1}{2}\Sigma^{-1}\end{pmatrix}.$$

The density can be written as an exponential family

$$\begin{aligned}h(x;\mu,\sigma) &= \frac{1}{\sqrt{(2\pi)^n \det\Sigma}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\\ &= (2\pi)^{n/2}\exp(\theta_1^T x + x^T\theta_2 x - \frac{1}{4}\theta_1^T\theta_2^{-1}\theta_1 - \frac{1}{2}\log\det(-2\theta_2))\\ &:= h(x;\theta).\end{aligned}$$

The terms inside the exponent form a concave function of θ *(see,e.g., the lecture notes of "Optimierung WiSe2021"), so*

$$E^Q[G(Z)^2\frac{g(Z)}{h(Z;\theta)}]$$

is a convex function of θ*.The sample average approximation of this leads to a convex "semidefinite programming" problem (*θ_1 *is a vector and* θ_2 *is a semidefinite matrix).*

Laplace method

The second moment can also be written as

$$\begin{aligned}
E^h\left[G(Z)\frac{g(Z)}{h(Z)}\right]^2 &= E^Q[G(Z)^2 e^{-Z^T\mu+\frac{1}{2}\mu^T\mu}] \\
&= E^Q[e^{2F(Z)-Z^T\mu+\frac{1}{2}\mu^T\mu}] \\
&= \frac{1}{(2\pi)^{n/2}}\int e^{2F(z)-z^T\mu+\frac{1}{2}\mu^T\mu-\frac{1}{2}z^Tz}dz
\end{aligned}$$

where $F(z) := \ln G(z)$.

To find a drift which reduces the variance, we now approximate

$$E^h\left[G(Z)\frac{g(Z)}{h(Z)}\right]^2 \sim Le^{\max_z\{2F(z)-z^T\mu+\frac{1}{2}\mu^T\mu-\frac{1}{2}z^Tz\}}$$

[this is heuristics based on "Laplace's method" which gives asymptotics for integrals of the form $\int e^{\frac{1}{\epsilon}F(z)}dz$ when $\epsilon \searrow 0$]. Thus we end up with

$$\min_\mu \max_z \{2F(z) - z^T\mu + \frac{1}{2}\mu^T\mu - \frac{1}{2}z^Tz\}.$$

Assuming that we can find the optimal μ (and z) as the saddle point (below, the first line is the partial derivative w.r.t. z and second w.r.t. x)

$$\begin{aligned}
2\nabla F(z) - \mu - z &= 0 \\
-z + \mu &= 0,
\end{aligned}$$

we end up with the optimality condition that μ solves

$$\nabla F(z) = z. \tag{7.1}$$

It is possible to show that this optimality condition is rigorous "asymptotic optimality condition" for finding the optimal change of drift for $e^{F(Z/\sqrt{\epsilon})/\epsilon}$ when $\epsilon \searrow 0$. We omit the details.

Importance sampling of Asian options with the Laplace method

We consider the Asian call option

$$c_{AC} = (\bar{S}_T - K)^+$$

where $T = 1$, $\bar{S}_T = \frac{1}{N}\sum_{n=1}^N S_{t_n}$, and $t_n = n/N$. Here, with a slight abuse of notation $S_n = S_{n\Delta t}$,

$$G(z) = f(\frac{1}{N}\sum_{n=1}^N S_n(z)),$$

where $f(y) = (y-K)^+$ and

$$S_n(z) = S_{n-1}(z)e^{(r-\frac{1}{2}\sigma^2)\Delta t+\sigma\sqrt{\Delta t}z_n}.$$

The optimality condition (7.1) becomes (when $\bar{S}_T(z) > K$)

$$z_n = \partial_{z_n}(\ln G)(z) = \frac{1}{NG(z)}\sigma\sqrt{\Delta t}\sum_{j=n}^{N} S_j(z) \quad n = 1, \dots N.$$

This means that

$$z_1 = \frac{\sigma\sqrt{\Delta t}[G(z)+K]}{G(z)}, \quad z_{n+1} = z_n - \frac{\sigma\sqrt{\Delta t}S_n}{NG(z)} \quad n = 1, \dots, N-1. \tag{7.2}$$

We leave it as an exercise to check that if we can find a scalar y such that $G(\mathbb{1}y) = y$, we get z_n recursively from the above equations with $G(z) = y$. Solving $G(\mathbb{1}y) = y$ amounts to finding the root of the one-dimensional equation

$$\frac{1}{N}\sum_{n=1}^{N} S_n(\mathbb{1}y) - K - y = 0.$$

Note here that the optimal drift vector $\mu = z$ is decreasing.

7.2 Exercises

Exercise 7.2.1. *Apply the importance sampling method to the European call option. Find numerically the optimal drift and compare the confidence intervals of the importance sampling method to the "naive Monte Carlo".*

Exercise 7.2.2. *How does the optimal drift depend on the model parameters? Compare the optimal drifts of European put and call options that are "deep out of the money", that is, for the put option, the initial asset price x is much larger than the strike K, and for the call option, vice versa.*

Exercise 7.2.3. *Apply the Monte Carlo method to estimate the price of the Asian call option. Plot the price as a function of the initial asset price x.*

7.3 Antithetic variates

The method of antithetic variates is based on the structure that the sampled (multi-dimensional) random variables Z satisfy that Z and $(-Z)$ are equal in distribution.

Again, for our applications, it suffices to assume that Z is a multi-dimensional standard normal which has this property. We define

$$Y := (G(Z) + G(-Z))/2$$

so that $EY = EG(Z)$ and we may perform Monte Carlo for Y instead to approximate $EG(Z)$. Implementation of this is extremely simple and we omit the detailed description of the algorithm.

To analyze the performance, note that

$$\begin{aligned}\operatorname{Var}(Y) &= \frac{1}{4}(\operatorname{Var}(G(Z)) + \operatorname{Var}(G(-Z)) + 2\operatorname{Cov}(G(Z), G(-Z)) \\ &= \frac{1}{2}\operatorname{Var}(G(Z)) + \frac{1}{2}\operatorname{Cov}(G(Z), G(-Z)),\end{aligned}$$

so for negative covariances the variance is reduced compared to the naive Monte Carlo. Of course, computationally, $Y^{(m)}$ is about twice as expensive to simulate. Note that Y and $\tilde{Y} := (G(Z) - G(-Z))/2$ are uncorrelated and that $Y + \tilde{Y} = G(Z)$, so

$$\operatorname{Var}(G(Z)) = \operatorname{Var}(Y) + \operatorname{Var}(\tilde{Y}).$$

As a rule of thumb, the more effective the method the "more linear" G.

7.4 Control variates

The idea behind control variates is to consider

$$Y(b) := G(Z) - b(X - EX)$$

where the expectation of the additional random variable X is known and $b \in \mathbb{R}$ is a parameter. We have $EY(b) = EG(Z)$, so we may do Monte Carlo for Y as soon as we can generate independent samples of (Z, X). In practice, $X = H(Z)$ for some function H, so generating the samples is not a problem.

To analyze the performance, we compute

$$\operatorname{Var}(Y(b)) = \operatorname{Var}(G(Z)) - 2b\operatorname{Cov}(G(Z), X) + b^2\operatorname{Var}(X),$$

so the variance is minimized with the choice

$$b^* = \frac{\operatorname{Cov}(G(Z), X)}{\operatorname{Var}(X)}$$

which gives (the right side is in terms of "correlation" $\rho_{G(Z),X} := \frac{\operatorname{Cov}(G(Z),X)}{\operatorname{std}(G(Z))\operatorname{std}(X)}$)

$$\operatorname{Var}(Y(b^*)) = \operatorname{Var}(G(Z))(1 - \frac{\operatorname{Cov}(G(Z))^2}{\operatorname{Var}(G(Z))\operatorname{Var}(X)}) = \operatorname{Var}G(Z)(1 - \rho^2_{G(Z),X}).$$

The more the variance is reduced the more correlated the control variate X is with $G(Z)$.

In practice, b is not known but it can be estimated from the sample covariances and variances. Before the actual Monte Carlo, we first generate a sample

$(Z^{(m)}, X^{(m)})_{m=1}^{\tilde{M}}$ (with size $\tilde{M}$) and use

$$\hat{b} = \frac{\sum_{m=1}^{\tilde{M}}[(G(Z^{(m)}) - \hat{\mu}^{(M)})(X^{(m)} - E[X])]}{\sum_{m=1}^{\tilde{M}}(X^{(m)} - E[X])^2}$$

as an approximation of the optimal parameter b.

Remark 7.2. *The method generalizes to sums of control variates*

$$Y = G(Z) - \sum_{i=1}^{K} b^i(X^i - E[X^i]).$$

Control variates in pricing

In our pricing application, two natural classes of control variates consists of options whose prices are explicitly known and of wealth processes of admissible portfolios.

Of options, the simplest choice is the forward (discounted price process) $X = \tilde{S}_T$ for which $E^Q\tilde{S}_T = S_0$, by the definition of Q. The Black Scholes pricing formulas, Theorems 4.4 and 4.5, allow us to use European calls and puts as well. Yet another choice is to use the quadratic option $X = S_T^2$ whose price was computed in Exercise 4.3.4. Explicit pricing formulas exist for a handful of other options as well (e.g., for barrier options), but these are not covered in these lecture notes.

Let z be an adapted process such that $z_{t-1}\Delta\tilde{S}_{t_n}$ is integrable for every n. Then

$$X_T = X_0 + \sum_{i=1}^{n} z_{t_{i-1}}\Delta\tilde{S}_{t_n}$$

is a Q-martingale with $EX_T = X_0$. Thus appropriate stochastic integrals (especially wealth processes) serve as control variates. When pricing exotic options, the whole path of S has to be simulated in the first place, so, in this case, the computational cost does not increase dramatically.

7.5 Exercises

Exercise 7.5.1. *Apply the importance sampling method to estimate the price of the Asian call option. Write the Matlab-code so that the importance sampler is easy to implement with varying drift vectors of the sampler. Use a "decreasing" drift vector to estimate the price of the Asian call option. Compare the confidence intervals of the importance sampling method to the "naive Monte Carlo". How would you change the sampler to price Asian put options? For the last question, no theoretical considerations are needed, just try to find a drift vector that performs better than "naive Monte Carlo".*

Exercise 7.5.2. *Apply the method of antithetic variates to estimate the price of the European put option. Compare the efficiency when the option is "deep in the money" and "deep out of the money".*

Exercise 7.5.3. *Apply the method of control variates to estimate the price of the European call option. Use the underlying stock as a control variate. Compare the efficiency when the option is "deep in the money" and "deep out of the money".*

Exercise 7.5.4. *Apply the method of control variates to estimate the price of the Asian call option. As a control variate, use the European call option with the same strike. Compare the efficiency when the option is "deep in the money" and "deep out of the money".*

8 Optimal stopping

We denote by $\mathcal{R}^1$ the space of continuous adapted processes R for which

$$\{R_\tau \mid \tau \in \mathcal{T}\}$$

is uniformly integrable. Such processes are called continuous processes of class (D). Let $R \in \mathcal{R}^1_+$ and consider the optimal stopping problem

$$\text{maximize} \quad ER_\tau \quad \text{over} \quad \tau \in \mathcal{T}.$$

We will first establish general existence and duality results. Later on, R takes the form $R_t = \phi(S_t)$ (e.g., American put option).

We first write the problem as

$$\text{maximize} \quad E\int Rdx \quad \text{over} \quad x \in \mathcal{C}_e,$$

where $\mathcal{C}_e := \{x \in FV \,|\, x_t \in \{0,1\}\}$. The equation $\tau(\omega) = \inf\{t \in \mathbb{R} \mid x_t(\omega) \geq 1\}$ gives a one-to-one correspondence between the elements of $\mathcal{T}$ and $\mathcal{C}_e$. Consider also the convex relaxation

$$\text{maximize} \quad E\int Rdx \quad \text{over} \quad x \in \mathcal{C},$$

where $\mathcal{C} := \{x \in FV_0^\infty \,|\, x \text{ increasing}, x_T \leq 1\}$ and FV_0^∞ is the set of adapted right continuous processes starting from zero with essentially bounded variation. Clearly, $\mathcal{C}_e \subset \mathcal{C}$ so the optimum value of optimal stopping is dominated by the optimum value of the relaxation. The elements of $\mathcal{C}$ are called *randomized stopping times.*

Recall that $x \in \mathcal{C}$ is an *extreme point* of $\mathcal{C}$ if it cannot be expressed as a convex combination of two points of $\mathcal{C}$ different from x.

Lemma 8.1. *The set $\mathcal{C}$ is convex, $\sigma(FV, \mathcal{R}^1)$-compact and $\mathcal{C}_e$ is the set of its extreme points.*

Proof. The set $\mathcal{C}$ is a closed convex subset of the unit ball that FV has as the dual of the Banach space $\mathcal{R}^1$ when $\mathcal{R}^1$ is equipped with the norm

$$\|R\|_{\mathcal{R}^1} := \sup_{\tau\in\mathcal{T}} E|R_\tau|.$$

The compactness thus follows from the Banach-Alaoglu theorem. It is easily shown that the elements of $\mathcal{C}_e$ are extreme points of $\mathcal{C}$. On the other hand, if $x \in \mathcal{C}\backslash\mathcal{C}_e$ there exists an $\bar{s} \in (0,1)$ such that the processes

$$x^1_t := \frac{1}{\bar{s}}[x_t \wedge \bar{s}] \quad \text{and} \quad x^2_t := \frac{1}{1-\bar{s}}[(x_t - \bar{s}) \vee 0]$$

are different elements of $\mathcal{C}$. Since $x = \bar{s}x^1 + (1-\bar{s})x^2$, it is not an extreme point of $\mathcal{C}$. $\square$

The following gives a dual expression for the optimum value as well as optimality conditions for the relaxed problem in terms of martingales that dominate the reward process R. We will denote the set of martingales of class (D) by $\mathcal{R}^1_m$.

We will use the following result from convex analysis.

Theorem 8.2. *Let $L : FV \times L^1 \to \mathbb{R}\cup\{\pm\infty\}$ be a concave-convex function such that $L(\cdot,x)$ is usc and $L(x,\cdot)$ is lsc for every $(x,y) \in FV \times L^1$, and, for some $x \in FV$, $\{y \in L^1 \mid L(x,y) > -\infty\}$ is uniformly integrable. Then there exists a saddle value*

$$\sup_{x\in FV}\inf_{y\in L^1} L(x,y) = \min_{y\in L^1}\sup_{x\in FV} L(x,y).$$

Moreover, $\bar{x}$ minimizes the left side and $\bar{y}$ maximizes the right side if and only if $(\bar{x},\bar{y})$ is a saddle point:

$$L(x,\bar{y}) \geq L(\bar{x},\bar{y}) \geq L(\bar{x},y) \quad \forall x \in FV, y \in L^1.$$

Theorem 8.3. *Optimal stopping time exists for every nonnegative $R \in \mathcal{R}^1$, the optimum value equals*

$$\inf\{EM_0 \,|\, M \in \mathcal{R}^1_m,\ R \leq M\},$$

where the infimum is attained. A stopping time τ is optimal if there exists $M \in \mathcal{R}^1_m$ with $M \geq R$ and $M_\tau = R_\tau$.

Proof. By Krein-Milman theorem, a continuous linear functional attains its supremum over a compact convex set at an extreme point of the set. The first claim thus follows from Lemma 8.1.

We use the fact that for any $R \in \mathcal{R}^1$, there is a nonadapted continuous process Z with $r := \sup_t |Z_t| \in L^1$ and $R_t = E[Z_t \mid \mathcal{F}_t]$. We note that the optimal value of the convex relaxation coincides with

$$\underset{x\in FV^\infty_0}{\text{maximize}} \quad E\left[\int R dx - r(x_{T+} - 1)^+\right] \quad \text{subject to} \quad dx \geq 0 \tag{8.1}$$

Indeed, for $x \in \mathcal{C}$ the two objectives coincide while if $x \in \mathcal{N}_0^\infty$ with $dx \geq 0$, then $\tilde{x} := x \wedge \mathbb{1}$ belongs to $\mathcal{C}$ and

$$\begin{aligned} E\int Rd\tilde{x} &= E[\int Rdx - \int Rd(x-\tilde{x})] \\ &= E[\int Rdx - \int Zd(x-\tilde{x})] \\ &\geq E[\int Rdx - r(x_{T+}-1)^+]. \end{aligned}$$

We define a concave-convex function on $FV \times L^1$ by

$$L(x,y) := \begin{cases} -\infty & \text{if } dx \not\geq 0, \\ E[+\int Rdx - yx_T + y] & \text{if } dx \geq 0 \text{ and } 0 \leq y \leq r, \\ +\infty & \text{otherwise,} \end{cases}$$

for which all the assumptions of Theorem 8.2 are satisfied (an exercise).

It is easy to verify that $\sup_y L(x,y)$ coincides with the objective in (8.1). On the other hand, we can write

$$L(x,y) = \begin{cases} -\infty & \text{if } dx \not\geq 0, \\ E[\int (R-M)dx + M_T] & \text{if } dx \not\geq 0 \text{ and } 0 \leq M_T \leq r, \\ +\infty & \text{otherwise,} \end{cases}$$

where $M_t = E[y \mid \mathcal{F}_t]$. Thus

$$\sup_{x \in FV_0^\infty} L(x,y) = \begin{cases} EM_0 & \text{if } 0 \leq M_T \leq r \text{ and } M - R \geq 0, \\ +\infty & \text{otherwise,} \end{cases}$$

As to the optimality conditions, the saddle-point condition in Theorem 8.2, $L(x,\bar{y}) \leq L(\bar{x},\bar{y}) \leq L(\bar{x},y)$ for all x and y, implies, first of all, that $d\bar{x} \geq 0$ and $\bar{M} - R \geq 0$. Then the first inequality means that $\int (R-M)d\bar{x} = 0$ and the second that

$$\begin{cases} \bar{y} = 0 & \text{if } \bar{x}_T < 1, \\ \bar{y} \in [0,r] & \text{if } \bar{x}_T = 1, \\ \bar{y} = r & \text{if } \bar{x}_T > 1 \end{cases}$$

almost surely. Thus $x \in \mathcal{C}$ is primal optimal if and only it satisfies the conditions given in the statement. For a stopping time τ, it is clear from nonnegativity of R that the corresponding x_τ can be chosen so that $x_\tau = 1$. Thus the optimality conditions reduce to the ones given in the statement. $\square$

Remark 8.4. *If the rewards process is not continuous, optimal stopping time does not exist in general. A simple example is provided by the reward process*

$$R_t := t\mathbb{1}_{[0,T')}$$

for some $T' \in (0,T)$. It is clear that the optimal value for the optimal stopping problem is T' (for $r \equiv 0$) but there is no stopping time τ for which $ER_\tau = T'$. In such cases, one has to relax the problem to allow for called "quasi stopping times".

8.1 American options

We continue in the setting of the Black Scholes model. The holder of an American option with a continuous reward process R and maturity $T > 0$ has the right to exercise the option at any time $t \leq T$ with payoff R_t. We set

$$q_t := \text{ess inf}\{y_t \in L^0_+(\mathcal{F}_t) \mid \exists\, \theta \in \mathcal{N}_A :\ X^{\theta,y_t} = y_t,\ X^{\theta,y_t}_u \geq R_u \text{ a.s. } \forall u \in [t,T]\},$$
$$p_t := \text{ess sup}\{y_t \in L^0_+(\mathcal{F}_t) \mid \exists\, \theta \in \mathcal{N}_A, \tau \in \mathcal{T}_t :\ X^{\theta,-y_t} = -y_t, X^{\theta,-y_t}_\tau \geq -R_\tau \text{ a.s.}\}.$$

The first is the *the superhedging price* for the seller while the latter is the *subhedging price* for the buyer. Note that these prices are not symmetric, since the holder (buyer) chooses the exercise time $\tau \in \mathcal{T}$ while the seller hedges against every possible action of the buyer.

We denote $\tilde{R}_t = \frac{R_t}{S^0_t}$, the *discounted reward process*. Note that p_t and q_t can be expressed in terms of discounted wealth and reward processes as well.

Theorem 8.5. *We have*

$$p_t \leq S^0_t \sup_{\tau \in \mathcal{T}_t} E^Q[\tilde{R}_\tau \mid \mathcal{F}_t] \leq q_t \quad a.s..$$

Proof. Let y_t, θ and τ be such that the condition in the definition of p_t is satisfied. Then

$$\tilde{X}^\theta_\tau = -\frac{y_t}{S^0_t} + \int_t^\tau \tilde{\theta}_s \sigma dB_s \geq -\tilde{R}_\tau \quad \text{a.s..}$$

By the supermartingale property of $\tilde{X}^\theta$ under Q (see the proof of Theorem 4.1),

$$-\frac{y_t}{S^0_t} \geq E^Q[\tilde{X}^\theta_\tau \mid \mathcal{F}_t] \geq E^Q[-\tilde{R}_\tau \mid \mathcal{F}_t] \quad \text{a.s.}$$

so we see that $y_t \leq S^0_t \sup_{\tau \in \mathcal{T}_t} E^Q[\tilde{R}_\tau \mid \mathcal{F}_t]$, which gives the first inequality in the statement.

Let y_t and θ be such that the condition in the definition of q_t, which implies that $X^{\theta,y_t}_\tau \geq R_\tau$ for $\tau \in \mathcal{T}_t$. Then, as above, $y_t \geq S^0_t E^Q[\tilde{R}_\tau \mid \mathcal{F}_t]$. Since $\tau \in \mathcal{T}_t$ was arbitrary, the second inequality in the statement holds. $\square$

When $p_0 = q_0$, we call

$$\sup_{\tau \in \mathcal{T}} E^Q \tilde{R}_\tau$$

the *risk-neutral price* of the American option with reward R.

Theorem 8.6. *Every American option with reward R such that $\tilde{R} \in \mathcal{R}^1_+(Q)$ has a unique risk-neutral price.*

Proof. We apply Theorem 8.3 (with $P = Q$). This gives a Q-martingale M and $\tau \in \mathcal{T}$ such that $M_t \geq \tilde{R}_t$ for all t, $M_\tau = \tilde{R}_\tau$ and

$$\sup_{\tau \in \mathcal{T}} E^Q[\tilde{R}_\tau] = E^Q[\tilde{R}_\tau] = E^Q[M_\tau].$$

Applying the martingale representation theorem (Theorem 2.21) to M, there exists, as in the proof Theorem 4.1, $\theta \in \mathcal{N}_A$ such that $\tilde{X}^{\theta,y_0}_t = M_t$, where $y_0 = E^Q[M_T]$. Then we see that $-\theta \in \mathcal{N}_A$ and $\tilde{X}^{-\theta,-y_0} = -M_\tau = R_\tau$ so that $p_0 = q_0$. □

8.2 Pricing American options with Monte Carlo

In this section, we apply the duality result, Theorem 8.3, to derive a Monte Carlo method to price American options. Another route would be to derive an "obstacle PDE" that the prices (as a function of time and state) have to satisfy and then apply finite difference methods similarly to as we did with the Black Scholes PDE in the case of Vanilla options.

Note that for any martingale $M \in \mathcal{R}^1_m$

$$\sup_{\tau \in \mathcal{T}} ER_\tau = \sup_{\tau \in \mathcal{T}} E(R_\tau + M_T - M_\tau) \leq E \sup_{t \in [0,T]} (R_t + M_T - M_t).$$

where the last expression is dominated by EM_0 if $R \leq M$. Thus,

$$\begin{aligned}
\sup_{\tau \in \mathcal{T}} ER_\tau &\leq \inf_{M \in \mathcal{R}^1_m} E \sup_{t \in [0,T]} (R_t + M_T - M_t) \\
&\leq \inf_{M \in \mathcal{R}^1_m} \{E \sup_{t \in [0,T]} (R_t + M_T - M_t) \,|\, R \leq M\} \\
&\leq \inf_{M \in \mathcal{R}^1_m} \{EM_0 \,|\, R \leq M\},
\end{aligned}$$

where, by Theorem 8.3, the last expression equals the first one. The optimum value of the stopping problem thus equals

$$\inf_{M \in \mathcal{R}^1_m} E \sup_{t \in [0,T]} (R_t + M_T - M_t). \tag{8.2}$$

This leads to numerical methods to derive upper bound to the optimal value of the optimal stopping problem. Note also that if Y is the Snell envelope of R (the smallest supermartingale that dominates R), then the martingale part M in the Doob–Meyer decomposition $Y = M - A$ is dual optimal.

8.3 Exercises

Exercise 8.3.1. *Consider the discounted American put option*

$$\tilde{R}_t = e^{-rt}(K - S_t)^+$$

in the Black Scholes model with $T = 0.5$*,* $r = 0.06$*,* $K = 100$*,* $\sigma = 0.4$*. Use* (8.2) *to estimate the price of the option for initial prices* $x = 80, 85, 90, 95, 100$*. Here, replace the set of martingales by multiples of* $M_t := e^{r(T-t)}\pi_P(t, S_t)$*, where* π_P *is the pricing functional of the corresponding European put with the same strike* K *(*M *is indeed a martingale by* (4.4)*; see also the proof Theorem 4.1).*

Note that the method gives upper bounds for the prices, whose approximations are (to check your solution) for $x = 80, 85, 90, 95, 100$*, respectively,* 21.6059*,* 18.0374*,* 14.9187*,* 12.2685*,* 9.9703*.*